Investigation of Security and Spectrum Management Issues in Cognitive Radio Aided by Machine Learning

Thesis by

Shaher Suleman Mousa Slehat

In Partial Fulfillment of the Requirements of the Requirements for the Degree of Doctor of Philosophy

University of Technology Sydney
Faculty of Engineering and Information Technology

Supervisor

Zenon Chaczko

Autumn, 2020

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Shaher Slehat declare that this thesis, is submitted in fulfilment of the requirements for the award of Phd, in the School of Electrical and Data Engineering/ Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: 10/01/2020

Acknowledgments

First and foremost, i would like to thank Allah the Al-Mighty for blessing me with the will, dream, and the resource to complete my PhD work.

My appreciation and special thanks are also due to my supervisor Dr. Zenon Chaczko for his invaluable support and guidance. He has indeed been an enthusiastic supporter of me work, providing a nearly unending stream of ideas. And providing me the independence to pursue my interests, and lending me the moral support, to clear the final hurdle. I am forever indebted to him for all his moral support and encouragement. And he always made me go this additional mile to resolve the different issues that lead to this work. i cherish the opportunity to see and learn from his experience and knowledge. His continual insights and patience with me are constantly appreciated.

I would also like to his opportunity to express my appreciations to my co-supervisor, Professor Robin Braun for supporting me throughout this work. Through this research, my grateful thanks are also due to the librarians of UTS main librarians who helped me in one way or

other both efficiently and courteously.

I also wish to thank all my friends and fellow participants who have supported and helped me over the last few years: Dr. Pakawat Pupatwibul and Dr. Anup Kale for their help in difference ways and from time to time.

Last but no the least, I wish to give special thanks to my family, Mr. Suleman Mousa Slehat, Ms. Seeta Helal , my brothers, and my sisters, for their massive support and all of the sacrifices that the have made on my behalf. My parents constantly gave me continual support and tried to provide me with the best education they can tolerate. They have been an important leading force to promote me behind this PhD research

Abstract

Cognitive Radio (CR) is an intelligent and adaptive radio and network technology that allows transceivers to sense available frequency spectrum and change its parameters, to switch to available channels(frequency bands) without interruption to other connected transceivers. It is primarily a technology to resolve spectrum scarcity problems using Dynamic Spectrum Access (DSA). The potential aspects and applications of Cognitive radio are far superior to DSA alone. CR abilities and CR reconfiguration abilities are essential components for electronic warfare (communications). It provides capabilities for developing and deploying advanced anti-jamming methods, by assisting in the development of advanced intelligent, self-reconfiguration methods to alleviate the effects of jamming.

This thesis examines the effects of jamming and other attacks on Cognitive Radio Networks and provides methods and processes to overcome those effects. Cognitive Radio architecture simulation was applied so that policies and their application correlate to Cognitive Radio jamming and anti-jamming issues. Simulation is employed for test-

ing Multi-Armed Bandit and machine learning strategies/solutions as shown by this thesis. The central part of the thesis is the mitigation of jamming outcomes on Cognitive Radio Networks by using proactive steps to increase communication robustness and contentiousness. The thesis utilizes game theory (i.e. the Multi-Armed Bandit problem) and protection using Machine Learning (ProML) design for analyzing jamming behavior on Cognitive Radio systems. MAB experiment show MAB approach is effective giants random attack, whereas, the proposed machine learning has its own merits to overcome constant and reactive jamming.

Contents

List of Figures

List of Tables

Nomenclature

16-QAM	16-Quaternary Amplitude Modulation
BPSK	Binary Phase Shift Keying
BSSI	Basic Service Set Identifiers
CPU	Central Processing Unit
CR	Cognitive Radio
DoS	Denial-of-Service
DSL	Digital Subscriber Lines
FPGA	Field Programmable Gate Array
HRRSS	High Relative Received Signal Strength
HWP	Hardwaer Platform
MAC	Media Access Control
MIMO	Multiple-Input Multiple-Output
MTS	Mobile Telecommunication Services
NCR	Network Cognitive Radio
NIICT	National Institute of Information and Communication and Technology
OA	Optimal Algorithm
OSA	Opportunistic Spectrum Access

ProML	Protection using Machine Learning
PU	Primary Users
PUE	Primary User Emulation
QoS	Quality of Service,
QP-SK	Quaternary Phase-Shift Keying
QPAM	Quaternary Phase Amplitude Modification
RF	Radio Frequency
RFU	Radio Ferquency Uint
RSSI	Received Signal Strength Indication
SDR	Software Defined Radio
SIP	Softwaer Infrastructure Platform
SNR	Signal-to-Noise Ratio
SPU	Signal Processing Unit
SUs	Secondary Users
TCP	Transmission Control Protocol
WLAN	Wireless Local Area Network
WNAN	Wireless Network After Next

1 Introduction

Bandwidth currently available to Cognitive Radio (CR) users is within the frequency range spectrum of 800 MHz to 6000 MHz. This spectrum is also utilized for cellular communications, broadband and other communication. This bandwidth span is from the very high-frequency bands (VHF) to the ultra-high frequency band UHF. Cellular and broadband communication, as well as other methods of mobile communication, are thought of as Mobile Telecommunication Services (MTS). In the early days radio transmission technologies development, the radio spectrum was largely uncontrolled and generally the radio spectrum was used inefficiently (Harada, 2008). Technologies such as Cognitive Radio were established for the prime purpose of bringing efficiencies to the use of the spectrum.

Cognitive Radio (CR) is an intelligent and adaptive radio and network technology. It allows transceivers to sense available channels within the spectrum, dynamically changing its parameters to switch to other available channels without interruption to other spectrum users (Karunambiga et al., 2015). Mobile Telecommunications Ser-

vices, in the vast majority, use channels within the spectrum which are suitable, and participates in new "solutions". Cognitive Radio is one of the solutions that can be used. Cognitive Radio is a technology, that as a system, is able to sense and provide complete awareness of its function, which is the organization of radio operating parameters, independently, in cooperation with other wired and wireless networks. CR provides a more predictable manner in applying available bandwidth opportunities in the spectrum band which in turn makes it an attractive, suitable, and predictable technology(Mitola Iii, 1999; Harada, 2008).

The concept of CR was first proposed by (Mitola Iii, 1999). In 2005, CR came into its own through a National Institute of Information and Communications and Technology (NIICT) project to improve Cognitive Radio technologies. The result of the project defined how software can update equipment Software Defined Radio (SDR), consisting of a Hardware Platform (HWP) and Software Infrastructure Platform (SIP) (Harada, 2008, 2005).

The principal part of the HWP is the Signal Processing Unit (SPU) which consists of a Field-Programmable Gate Array. The HWP consists of a "signal processing unit" (SPU) which consists of the "Field-Programmable Gate Array" (FPGA), a Central Processing

Unit (CPU), a multi-band antenna which supports the UHF (400 MHz to the 5 GHz band) and a multi-band RF Unit (RFU) which supports the 5GHz band.

The SIP consists of different applications which control spectrum sensing, reconfiguration of transmission parameters and overall sensing for communications opportunities, which improves the overall operating behaviors the NIICT project united Cognitive Radio technologies by enabling sensing "sign levels" above the 400MHz-6GHz bands and by selecting the model structure (prototype) using software applications to check for connection opportunities within the spectrum (Harada, 2008; Ahmed, 2010).

The major challenge is the availability of channels in the spectrum, where the different frequency bands (channels) are used by various applications and protocols, which may or may not interact with each other. Cognitive Radio meets this challenge by sensing the environments parameters and making decisions based on collecting information on discrete, and available frequency bands (channels) (Jinlong et al., 2015). While, the connotation of Cognitive Radio, as submitted by Mitola Iii, 1999, is to assist in resolving availability issues in the spectrum by selecting "the best" frequency bands (channels) from others that are available, it also assists by allowing Secondary Users

(SU) of unlicensed domains and Primary Users (PU), to access any unused channels within the licensed domains the (PU) spectrum. It assists in resolving any issues with channels, such as potential conflicts thus providing smooth connectivity and seamless services. Also, Cognitive Radio reinforces spectrum capacity by avoiding collisions and wastage (Rawat et al., 2016). Advanced Cognitive Radio has been developed over time and includes categories such as; SDR, Cognitive Radio (CR) and Network Cognitive Radio (NCR).

Cognitive Radio collects information about its frequency environment using past experiences. It senses parameters by combining and adapting information sensed from its neighbors. Cognitive Radio includes two kinds of users; the Primary User (PU), who is licensed to allow unfettered access to the spectrum and Secondary Users (SUs), who accesses unused, available frequency bands of the spectrum's unused channels. When the CR senses available frequency bands ahead of their use by the PU it is considered to be a " smart system"- which adapts itself and provides a credible smooth connection to the PU.

Research has assisted the design and architecture of Cognitive Radio and promoted its development. The research by Akyildiz et al., 2008; Fragkiadakis et al., 2013; Wang et al., 2010a research on Cognitive Radio includes developing the architectural design of Cognitive Ra-

dio focusing on spectrum management problems, security attacks and other challenges which may affect the way Cognitive Radio works.

Security attacks include radio "jamming" and other attacks. Different types of jamming attacks and anti-jamming in the wireless networks were introduced by (ÇakiroÇğlu and Özcerit, 2008; Grover et al., 2014; Wang and Wyglinski, 2011 focused into research jamming detection and reliable countermeasures. Bahl et al., 2004 research introduced the effects of jamming on Cognitive Radio. Their research included effective countermeasures such as multi-channel hopping in the spectrum (wireless networks) leading to upgrading the capability of Cognitive Radio wireless networks.

Wireless networks, generally, are sensitive to jamming attacks (Tomić and McCann, 2017). Varieties of attacks include bypassing the (MAC) layer protocol (lower sub-component of the data link layer, 7 layer OSI communication model) or by transmitting interference RF signals (Braithwaite, 2017). However generally, jamming is indicated as a "purposed" attack on users of a wireless network (i.e. Cognitive Radio). Jamming is a sharp Denial-of-Service (DoS) attack against a wireless network (Manogna and Naik, 2014).

My research focuses on jamming attacks within Cognitive Radio Networks, utilizing different anti-jamming strategies including variations

of the Multi Armed Bandit problem of which the Optimal Algorithm (OA) is an example.

1.1 Research Gap in Anti-Jamming Strategies

While significant research has been made into Cognitive Radio and its threats (jamming attacks), its performance, until now, is far from complete, even under ideal circumstances, because of "geographical differences" (different physical locations) and random interference from Jamming Attacks. Despite this, the general performance of the system of the Cognitive Radio Network is the sum of the individual users of the system. The main problem affecting the reliance on users, is the potential "hogging" of frequency bands by PUs, who collectively impede the individual systems performance for SUs. The correlation of Jamming attacks with system performance is under-researched. It shows total system collapse under Jamming Attacks (JAs), but not the degree of system degradation to PUs, and to a leaser extent, SUs, which impedes the Quality of Service (QoS) to all users (Zhou et al., 2018).

1.2 Motivations and Research Problem

Extensive research has been undertaken to investigate the Cognitive Radio Network and its many related issues and problems. Most of these have focused on analysis of various (often limited) scenarios of Cognitive Radio and Cognitive Radio based Network usage. The Cognitive Radio Network protection mechanism against Jamming Attacks is also a major field of research investigation.

Security and reliability issues of Cognitive Radio receive a high level of attention from the research community as it is perceived as a key enabler of fifth generation (and beyond) cellular network technologies, for Opportunistic Spectrum Access (Braithwaite, 2017). The main motivation of my research is to determine viable solutions for jamming attacks on Cognitive Radio Networks.

1.3 Research Significance

Large technological leaps are required to fill the gaps between the required reliability of current networks and the capability of technology that is currently designed to mitigate problems related to sharing a contested frequency spectrum. This research pays special attention to jamming of Cognitive Radio Networks. The significance of this work

is to further explore and discover the nature of jamming threats and look for more effective solutions. The chief aim is to provide Cognitive Radio with innovative algorithmic and associated methodological solutions characterizing better anti-jamming features than other competitive approaches.

1.4 Objective and Aims of Research

My research is based on the following objectives:

- To provide solutions protecting the Cognitive Radio Network, from jamming attacks.
- Designing an effective scheme for implementing these solutions to cover as much as possible the problem of jamming attacks in Cognitive Radio.
- Enhance the network Quality of Service (QoS) for Cognitive Radio Network users, both PUs and SUs.

1.5 Research Hypothesis

"Enhancing the current multi-arm bandit gaming approach by using dynamic channel selection methods which reduces the chances of jamming attacks, and improves channel availability in a Cognitive Radio Network".

1.6 Research Question

Is it possible to develop an efficient and reliable method to solve problems caused by Jamming Attacks on Cognitive Radio Networks, using solutions that avoid spectrum sensing?

1.7 Research Contribution

The main contribution of my research is summarized as follows:

- A comprehensive overview of security problems related to Cognitive Radio and Cognitive Radio Networks.
- Current Cognitive Radio technologies often introduce loopholes which enable jamming attacks; variations of the Multi-Armed Bandit problem are deployed to counter jamming attacks. In this work several other inventive methodologies introduce the experimental work which is the main contribution of this research.

1.8 Structure of Thesis

The structure of the thesis is shown in the table below and consists of two parts (see tables Structure of Thesis part 1 and part 2). Each part is composed of a number of chapters that present all relevant concepts

needed to encapsulate the topics discussed, including a bibliography. The thesis is organized as follows:

Part 1: Theory, Concepts and State of the Art Methods

- Chapter one provides an overview of my research work;
- Chapter two includes a literature review on applications of Cognitive Radio and security topics;

  * A brief overview of the background literature, classified according to key aspects and the recent state of Cognitive Radio Networks (CRNs).
  * A view of where current Cognitive Radio technology is now positioned.
  * An overview of literature on smart jamming and anti-jamming systems using Cognitive Radio applications.
  * An overview of key security problems involving Cognitive Radio and Cognitive Radio Networks.
  * Security inherited by Cognitive Radio and legacy wireless Cognitive Radio systems.
  * A review of related security problems and jamming attack problems in Cognitive Radio Networks.

    * Other research that has used the multi-armed bandit problem.

- Chapter three

    * An overview of game theory, as applied to this research, and an examination of strategies in the multi armed bandit problem used in its various scenarios.

Part 2: Research Work and Experimental Validation

- Chapter four

    * Proposed methods and strategies in deploying multi-armed bandit technologies.

    * Solutions to the problem to be addressed.

- Chapter five

    * Empirical measurements of proposed methods and strategies

    * Aspects of proposed methods as evaluated through outcomes and validation.

- Chapter six

    * Case studies utilizing the proposed methods and strategies.

    * Research proposal and related applications based on how the Cognitive Radio environment is applied to designed experiments.

- Chapter seven

  * Conclusion, outcomes and future research work on the topic.

Table 1.1: Structure of Thesis Part 1

Part 1: Theory, Concepts and State of the Art Methods			
Chapter Number	**Introduction**	**State of the Art Methods**	**Mathematical**
Chapter 1	*Introduction:* Thesis topic Objective Aims Contribution Outline		
Chapter 2		*Literature Review:* State of the Art	
Chapter 3			*Theoretical Apparatus:* Game Theory Multi-Armed Bandits Upper Confidence Bound Kullback-Leiber Bound Thompson Sampling

Table 1.2: Structure of Thesis Part 2

Part 2: Research Work and Experimental Validation				
Chapter Number	**Methodology**	**Experimental Results**	**Action Research**	**Conclusion**
Chapter 4	*Methodology:* Evolution Modification Modulation			
Chapter 5		*Experimental Results:* Evaluation Methods		
Chapter 6			*Action Research:* Cognitive Radio Environment	
Chapter 7				*Conclusion:* summarizes commentary Future Work Expansion

2 Literature Review

This chapter includes concepts associated with Cognitive Radio and Software Defined Radio technologies, discussing their progress from their inception until now. This chapter also includes discussion on applications associated with Cognitive Radio, in addition to tacit empowerment technologies. Finally, a review of latest developments in Cognitive Radio jamming and anti-jamming practices, is included.

2.1 Background of Cognitive Radio

Cognitive Radio (CR) is a technology that was developed as a model for supporting frequency spectrum access (Liang et al., 2011). CR is the most important technology for wireless communication networks to cognitively connect radio nodes. Cognitive Radio was defined by (Mitola Iii, 1999) in his seminal work: as a radio or system that senses, and is aware of, its operational environment and can dynamically and independently adjust its radio operating factors on that account". In general, Cognitive Radio is a radio or system that can sense its elec-

tromagnetic surrounds, directly and independently regulating its radio operating parameters to adjust its system's operation, so as to maximize throughput, mitigate interference, provide smooth interoperability and access unimportant markers" (Mitola Iii, 1999). This definition of Cognitive Radio has two key characteristic features that differentiate it from a traditional radio: the cognition capability and re-configurability.

Figure 2.1 shows the characteristic feature of Cognitive Radio, conceptually interacting with the surrounding broadcast environment. This design is referred to as the cognition cycle and is repeatedly run by each Cognitive Radio to determine spectral opportunities, make plans to adapt itself and determine better opportunities (Khattab et al.; Mitola Iii; Mitola).

Other definitions include:

- Cognitive Radio (CR) is an adaptive, smart radio and network technology that can mechanically discover available channels in a wireless spectrum and alter transmission parameters, therefore enabling more communications to run at the same time and also ameliorate radio-operating behavior. Cognitive Radio uses comprehensive adaptive radio technologies and at the same time pro-

vides a communications system watch (justifies its own performance). Software Defined Radio (SDR) is where conventional hardware components, including mixers, amplifiers and modulators have been replaced with intelligent software (Khattab et al., 2013).

- Cognitive Radio combines machine understanding software into wireless system radio nodes and networks. Radios today are working toward the realization of cognitive information access and users demeanor. This is not in just sensing the Radio frequency (RF) spectrum, but also in becoming aware and interpreting the users in their surroundings through computer vision, speech recognition, and language recognition (Mitola, 2005).

- Cognitive Radio technology is an active emerging cognitive wireless system that can find an available radio frequency in its neighborhood and connect nodes so as to provide a user with the best possible service. One critical capability of Cognitive Radio is its ability to seamlessly shift from one band of the radio spectrum that it is using, to another available one, to complete a communication channel, especially during an emergency (De Nardis and Holland, 2014).

- Cognitive Radio is a technology which enables the system to ob-

tain knowledge of its own operational and geographical environment; develop policies on its internal situation and to dynamically and independently modify its operational parameters and protocols based on acquired knowledge, in order to realize its objective which is previously determined and to learn from the acquired information (Nguyen et al., 2012).

- Cognitive Radio is a type of wireless device which sends and receives signals in a selected Radio Frequency (FR) zone of the electromagnetic spectrum in order to assist the transfer of data. Radio exists in many items such as computers, mobile phones, televisions, car door openers, and motor vehicles (Group et al., 2007).
- Cognitive Radio is a system that is capable of sending and receiving signals which can discover alternative channels in its surroundings and be able to adapt (Chaczko et al., 2010).

2.2 How Does Cognitive Radio Work?

A particular frequency "transmits" when you select your favorite station on, say, an AM/FM radio. The antenna circuit "tunes" to select the station by detecting frequency signals from the ether, and takes

many samples per second (as per the Nyquist-Shannon sampling theorem). Once the station is selected, switching on a receiver will automatically enable the sampling of data transmitted on that frequency (band), which is decoded into "useful" information (audio, vision, digital data) as determined by the user. Sometimes other signals interfere with the reception, particularly from nearby high powered Rf signal sources. This is often called "jamming".

2.3 Advantages of Cognitive Radio

Cognitive Radio technology assists users with individual needs on a large scale, as per:

- New spectrums are determined and employed automatically.
- Different network standards are interpreted and recognized automatically.
- Automatically developing methods to reduce interference.

2.4 Cognition Capability of a Cognitive Radio

The cognition aspect of Cognitive Radio is the ability of a CR transceiver to sense its surrounding radio environment, and analyze acquired data to determine the best actions as to which spectrum

bands to use, transmission strategies to be adopted, including cognition ability, for continuous monitoring of the dynamically and changing environment surrounding the radio, in order to set a suitable communication plane (Khattab et al., 2013). The cycle of Cognitive Radio cognition is shown in Figure 2.1, which defines Spectrum Sensing, Analysis, Decisions, Adaptation. The three major ingredients of the cognition cycle can be described in the following sub-sections:

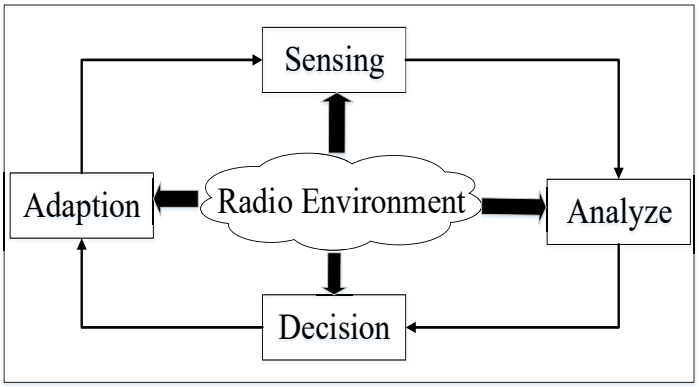


Figure 2.1: Cycle of Cognitive Radio, Adapted from (Khattab et al., 2013)

2.4.1 Spectrum Sensing

Spectrum Sensing is the most critical aspect of Cognitive Radio in the sense that it includes a capability to sense, learn, measure and provide awareness of the frequency spectrum used in its environment. See Figure 2.2.

Spectrum sensing is the highest and most important part of Cognitive Radio, as it provides awareness of the spectrum used in areas of the

user's environment (Yücek and Arslan, 2009). Cognitive Radio is required to decide, in real time, which frequency band is used to sense the detection time, the data acquisition cycle and possibly what is the estimated signal strength. Spectrum information must be sufficient for Cognitive Radio to reach correct conclusions to provide accurate information on the user environment. It must also provide rapid spectrum sensing to track temporal changes in the radio environment.

These requirements of spectrum sensing, set challenging conditions for hardware implementation in the Cognitive Radio environment. Bandwidth scale, processing power, Radio Frequency (RF) circuits and other current spectrum sensing techniques, all rely on the discovery of necessary transmission activities. These schemes are classified as: matching filtering detection; energy and detection features; and interference with temperature measurement detection (Khattab et al., 2013). Hence, spectrum sensing is the main function of a Cognitive Radio Network and therefore assists spectrum administration by layering protocols to transmit information about the spectrum.

2.4.2 Spectrum Analysis

Spectrum analysis deduces spectral opportunities in the surrounding radio environment according to sensed radio environment parameters. A spectral opportunity is traditionally defined as "a band of frequen-

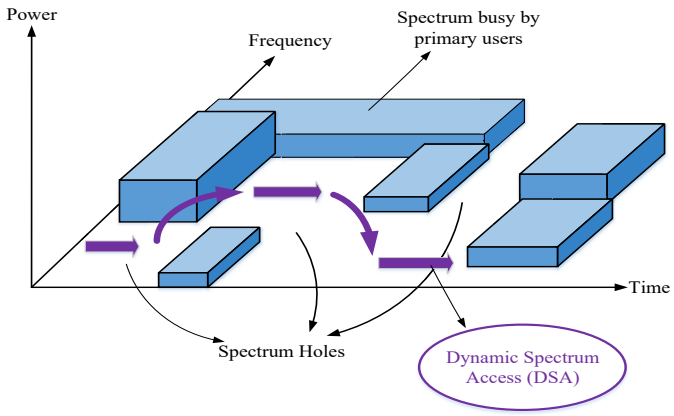


Figure 2.2: Spectrum Holes, adapted from (Yücek and Arslan, 2009)

cies unused by the primary user of that band, at a particular time, in a particular geographic area"(Kolodzy et al., 2001). Nevertheless, the definition is only sufficient to cover just three dimensions of the spectrum space which include: frequency, time, and space. However, it can exploit another dimension of the spectrum, the coding dimension, that employs distributed spectrum coding techniques for spectral opportunities in the spectrum band and which are actually used by its licensed users (PUs) .

Likewise, issues such as the "angle dimension" provide spectral opportunities due to the use of radial formation, to grant Cognitive Radio users the ability to transmit simultaneously on a currently utilized band. As well, contemporary developments in radio transmission techniques, such as the use of Multiple-Input Multiple-Output (MIMO) (Gaur et al., 2004) at the physical layer (layer 1 of the 7-layer OSI

Interconnection model), can create new dimensions of a spectral opportunity of a usable frequency band. For example, in monitoring a communication stream, an antenna can be used to provide Cognitive Radio users (PUs and SUs) an ability to communicate with each other and other licensed users (PUs) without interrupting the performance of legal users (PUs and SUs) (Khattab et al., 2013; Gaur et al., 2004).

The definition of spectral opportunity could be then considered as: “a theoretical hyperspace taken by radio signals, that has dimensions of location, the angle of arrival, frequency, and time" (Chaczko et al., 2010; Khattab et al., 2013; Yücek and Arslan, 2009).

2.4.3 Spectrum Decisions

The main role of a spectrum decision during the cognition cycle of an operational Cognitive Radio, is to decide on which group of transmission procedures to be accepted, based on results of spectrum sensing and analysis. Cognitive Radio uses information collected on spectrum bands, to identify spectral opportunities and to determine how a radio transceiver uses transmissions in those frequency bands. Transceiver parameters, to be decided on, rely on implicit transceiver architecture.

By sensing spectrum information and transceiver architecture, a Cognitive Radio defines the values of parameters to be configured for an

available transmission. Moreover, a spectrum decision includes co-undertaking spectrum selection and path structure. The "activity set" contains data about which spectrum is more suitable for transmission, the most recent time a transmission must start on a defined band, its transmission power, modulation rate, any spread spectrum hopping scheme, angle of a newcomer for directional transmissions, and the number and identity of antennas to be used (Akyildiz et al., 2009; Khattab et al., 2013).

2.4.4 Reconfigurability of Cognitive Radio

Figure 2.1 illustrates a second feature differentiating a Cognitive Radio from a conventional one, by re-configuring its ability to check transceiver parameters directly on the basis of an assessment of the radio environment. Cognitive Radio has great pliability, which is demonstrated by its capacity to reconfigure transmission parameters including the transmission rate and power. Cognitive Radio must be able to exploit opportunities of spectral availability in a wide spectrum range. Cognitive Radio must specify the bandwidth in which it travels, in order to adapt to different sizes of its spectrum. Moreover, Cognitive Radio should not be restricted to a particular communication protocol, and must also determine appropriate communication protocol to be used for different spectral opportunities (available frequency bands),

in its environment.

(Mitola Iii, 1999) developed the notion of Cognitive Radio software. In the Cognitive Radio environment the ideal is to implement radios with compositional abilities and smoother implementation of parameters via software applications. Cognitive Radio was originally created as a "software radio environment" having the capacity of being self-aware, over extended periods. Software radios cannot achieve data rates needed by wireless services because software and hardware platforms can be at "gridlock". This motivated research to develop a fast, multi-gigahertz, Cognitive Radio transceiver with smoother flexibility, provided by inexpensive, software applications (Khattab et al., 2013).

2.4.5 Spectrum Mobility

(Mitola Iii, 1999) Defined re-configurability of a Cognitive Radio transceiver: by spectrum mobility functions, and the change of spectrum mobility in a Cognitive Radio terminal, to maintain a smooth wireless connection. A radio terminal should be able to change to a new frequency band upon the arrival of a primary user (PU) of the spectrum and if that channel deteriorates, to substitute existing available channels (generally a Cognitive Radio user leaves the spectrum and resumes communication in another part of the spectrum). Radio spectrum mobility is the Cognitive Radio function which explores

obtainable cognitive spectrum opportunities in one or many usable frequency bands.

The association of spectrum mobility with a hands-off mechanism of transmission, assures the transmission spectrum of the new frequency band, without breaking it. Spectrum mobility assists the hands-off mechanism to expose link failures and to alter communication paths to new spectrum bands for connection and communication between Cognitive Radio Networks. With this cognition, Cognitive Radio has a significant impact on lower layers of the 7-layer OSI communication model, in the Cognitive Radio Network - i.e. the physical and transport layer access, while mobility spectrum and hand-off affects the session, presentation and application layers.

Spectrum mobility schemes must guarantee smoother and rapid transition frequency protocols, by modifying parameters accordingly and ignore latency, which may impact on communication protocols such as TCP and IP. Although mobility based hands-off operations have been verified in the context of cellular networks in general, there must be real cooperation between components of the Cognitive Radio life cycle in order to cope with obstacles from a restricted cognition network (Mitola, 2000; Khattab et al., 2013).

2.5 Cognitive Radio Spectrum Sensing

Because Cognitive Radio is used in many applications, spectrum sensing becomes very important, as Cognitive Radio's practical purposes are being utilized to present a technique of utilizing spectrum sensing more effectively. Spectrum sensing is the "password" for many software applications. The strength of any Cognitive Radio system is to access additional parts of the radio spectrum, and observe the spectrum to ensure it doesn't cause any unnecessary interference to devices that rely entirely on spectrum sensing of items in the system/spectrum.

2.5.1 Cognitive Radio Spectrum Sensing Basics

Generally, Cognitive Radio systems and other radio systems coexist within the same spectrum, without causing unnecessary interference, when sensing and using the spectrum. Any Cognitive Radio must consider:

- **Uninterrupted Spectrum Sensing**: The radio should uninterruptedly sense spectrum usage. Usually Cognitive Radios use the spectrum on the basis of non-interference for primary users (PUs).

- **Observe for Alternate Empty Spectrum**: The radio must determine any obtainable alternate spectrum to which it can switch a secondary user as needed, in the instance when a primary user requires the spectrum being used.
- **Observe Kinds of Transmission**: It is very important for Cognitive Radio to sense the kinds of transmissions made, and should be capable of defining the kind of transmission utilized by a primary user, so that fake transmissions and other similar interferences can be detected and disregarded.

2.5.2 Kinds of Cognitive Radio Spectrum Sensing

There are number of methods by which Cognitive Radio preforms spectrum sensing. Methods by which Cognitive Radio spectrum sensing can be made, fall into one of two classes:

- **Non- Cooperative Spectrum Sensing**: Occurs when a Cognitive Radio operates on its own. The Cognitive Radio operational format is based on signals which it detects and configuration information held in its activity set.
- **Cooperative Spectrum Sensing**: Occurs when a number

of cooperating radios throughout the Cognitive Radio Network share data. Usually, a central station receives reports on signals transmitted from different radios in the network and adapts the network to suit.

Generally, Cognitive Radio cooperation decreases the problems of interference, whenever a Cognitive Radio can not "hear" a Primary User because of problems like "shadowing" from another Primary or Secondary User.

2.6 Definition of Cooperative Spectrum Sensing

Cooperative Cognitive Radio spectrum sensing occurs when a network of Cognitive Radio participates in sensing available channel information from radios operating within the network. This provides a useful picture of spectrum usage in the network where the Cognitive Radio operates. There are two approaches to cooperative spectrum sensing :

- **Centralized approach**: A main node or radio collects all information from other operational nodes within the network, which

it analyzes and decides which frequency bands or channel can be used.

- **Distributed approach**: All nodes share the information.

2.7 Application of Cognitive Radio

Cognitive Radio has many application capabilities, including:

- Non-real time applications, such as mobile internet.
- Wireless networking such as "hot-spot" (Toolkit, 2010).
- Centralized multimedia networking (distribution networking), and
- High performance communication (Toolkit, 2010).

2.8 Architecture of the Cognitive Radio Network

A Cognitive Radio Network is composed of Primary Radio Networks operating in similar geographical regions as each other and include secondary networks operating in the same area. The primary radio network is licensed to operate in a "light" spectrum band (Khattab et al., 2013; Yücek and Arslan, 2009). A primary radio network has centralized services (infrastructure) and ad-hoc radio sectors in its immediate environment. PUs (Primary users) have primacy on spectrum access and operate as the sole users of their licensed spectrum. PUs

do not support any communication with secondary networks. Figure 2.3 shows the Cognitive Radio Network architecture.

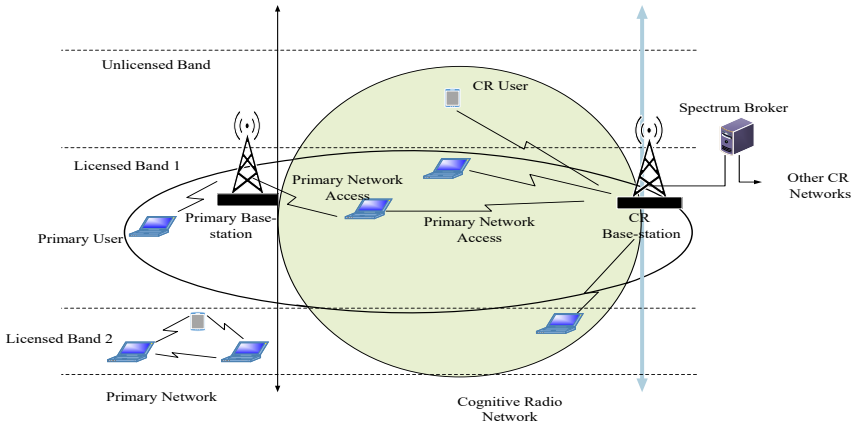


Figure 2.3: Cognitive Radio Network Architecture

The Cognitive Radio Network is a non-parasitical network. By that is meant, that the PU transmission in the primary radio network is not affected by secondary users. Accordingly, the Primary Networks (PNs) determine "upper bounds" on activities in their licensed bands, typically in terms of higher power levels, to ensure guaranteed performance levels to their authorized users.

Furthermore, secondary users (unlicensed) operate in a predefined band. The spectrum is accessed in an opportunistic way, that permits secondary users to opportunistically access the available, geographically complete spectrum. Cognitive Radio users can exploit an unlicensed spectrum. This indicates a spectrum heterogeneity of CRNs. When operating in a licensed band, transmissions must comply with

restrictions imposed by its licensed owner. The CRN can be either a centralized infrastructure network or a distributed one (Khattab et al., 2013; Yücek and Arslan, 2009). See figures 2.4 and 2.5.

2.8.1 Primary Network

A primary network includes licenses for particular radio frequency bands, and consists of networks such as WiMAX, CDMA, Cellular network, and broadcast networks (TV). A primary network includes the:

- Primary User (PU): A user who has highest priority or inheritance rights to the usage of a special part of the spectrum.
- Primary base station: A fixed base station, which is not a traditional Cognitive Radio node within a Cognitive Radio network. Therefore, it requires modification to access the licensed spectrum in the primary network.
- Secondary User (SU): A user who has the lowest priority and takes advantage of the spectrum in such a manner so as to cause no interference to primary users.
- Spectrum Sensing: Spectrum sensing gains awareness about spectrum uses and the presence of primary users in its geographical

region.

2.8.2 Cognitive Radio base station

The Cognitive Radio base station is a basic component, of constant structure, that facilitates all connections for CR users without regard to licensed channels in the spectrum. CR users have the ability to access other networks via the base-station and can be thought of as a "bridge" from one network to another.

2.8.3 Cognitive Radio user

Cognitive Radio users can be either a primary or secondary user. The secondary user is one who has the lowest priority and who takes advantage of the usable spectrum in such a manner that does not cause interference to primary users.

2.8.4 Spectrum broken

Spectrum broken is when the main network shares spectrum resources between many Cognitive Radio Networks. Spectrum broken enables connections to networks, such as star networks and can work as a centralized server that provides information about spectrum resources of various networks. Within the context of spectrum broken the major

components include:

- **Primary User**: They have a license to operate in a particular band of the spectrum. This can only be controlled from the base station and cannot be affected by unlicensed users.
- **Primary Base-Station**: A fixed network that has a licensed spectrum and has no ability to share the spectrum with non-licensed users, maybe needs a base-station to obtain legacy and Cognitive Radio protocols by accessing network information from primary users.
- **Cognitive Radio User**: Cognitive Radio user capabilities include: "spectrum sensing, spectrum decision, spectrum hand-off and Cognitive Radio MAC/routing/transport protocols" (Akyildiz et al., 2006). The Cognitive Radio user should have abilities to communicate not only with the base station but also with other CR users. Some CR users do not have a spectrum license. Therefore, spectrum access is only permitted by opportunistic use.

The CRN architecture permits three different access types across heterogeneous networks, including:

- **Cognitive Radio Network Access**: CR users have access to the base-station for licensed and unlicensed spectrum bands. Since all communications occur within the Cognitive Radio Network, the average access scheme is independent of the primary network.
- **Cognitive Radio Ad Hoc Access**: CR users communicate with other Cognitive Radio users over an ad hoc connection to licensed and unlicensed spectrum bands. Cognitive Radio users also have average access to available technology.
- **Primary Network Access**: CR user has access to the primary base-station over the licensed band, if the primary network permits. Unlike other kinds of access, Cognitive Radio users must support access technology to the primary network. In addition, the primary base-station should support Cognitive Radio abilities (Akyildiz et al., 2006, 2008).

2.8.5 Infrastructure Based Network (Centralized Cognitive Radio Networks)

Centralized Cognitive Radio Networks are based on an infrastructure that controls all radio stations particularly the transmission activities of secondary radio users. Figure 2.4 illustrates the centralized infras-

tructure based network.

The Cognitive Radio Network controls secondary transmissions through licensed and unlicensed bands, by aggregating spectrum data sensed from the network users. By using the aggregation data, the base station makes spectrum access decisions for all network nodes. The IEEE 802.22 networks model is the best example of a centralized CR and is its first universal standard. The IEEE 802.22 model works well with point-to-multi point communication via new television (TV) bands under base terminal control, utilizing a Cognitive Radio users framework and adaptive centralized spectrum databases within a 33 km radius.

There are other examples of the Cognitive Radio Networks like the two: “European Dynamic Radio for Internet Protocol (IP) services in a Vehicular Environment” and “Spectrum Efficient Uni-and Multi-cast Services Over Dynamic Radio Network in Vehicular Environments”. “These have centralized structures which regulate dynamic exploitation of temporary space of spectral opportunities” (Networks, 2014).

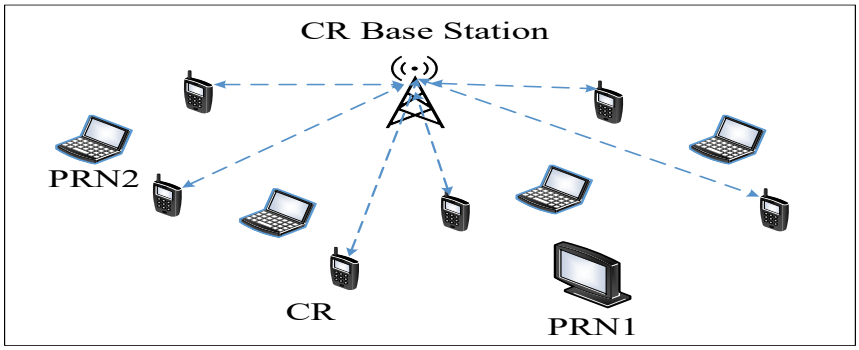


Figure 2.4: Centralized based Cognitive Radio, adapted from (Khattab et al, 2013)

2.8.6 Ad-hoc Network (Distributed Cognitive Radio)

Cognitive Radio Networks, as well as other nodes (not necessarily operational at that moment), communicate with each other by ad-hoc, point-to-point, Internet communications within licensed or unlicensed frequency bands. Cognitive Radio nodes apportionment in the network, coordinate their spectrum access decisions to share spectrum usage opportunities. A universal technique like full "network synchronization" is necessary for spectrum access coordination. Apportionment cooperation and other communication mechanisms are utilized to upgrade the network connection performance. Figure 2.5 shows the distribution of Ad-hoc Cognitive Radio Network. A result of mitigating infrastructure costs, is that infrastructure-less Cognitive Radio Networks impose complex network intricacy by lack of centralized control.

Examples of ad-hoc networks include the Cognitive Radio Networks (CRN) environment, "Peer-to-peer mode of DARPA's neXt Generation (XG) dynamic access network" (Ramanathan and Partridge, 2005). "DARPA's Wireless Network After Next (WNAN) military test bed" (Khattab et al., 2013) , and "Cognitive Radio approach for usage of Virtual Unlicensed bands (CORVUS)" (Brodersen et al., 2004).

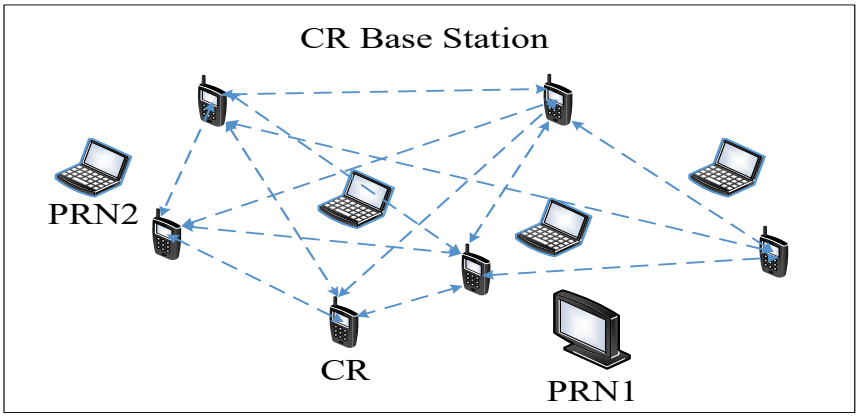


Figure 2.5: Distribution of Ad hoc Cognitive Radio, adapted from (Khattab et al., 2013)

2.8.7 Mesh Architecture

Mesh architecture is a compound of infrastructure and ad-hoc networks. Devices are connected to base stations via neighboring nodes and packets are directed and sent by base stations. For example, a network connection is distributed between wireless mesh nodes connected with each other, primarily to share network communication

over a wide area (Chen et al., 2008) as per figure 2.6.

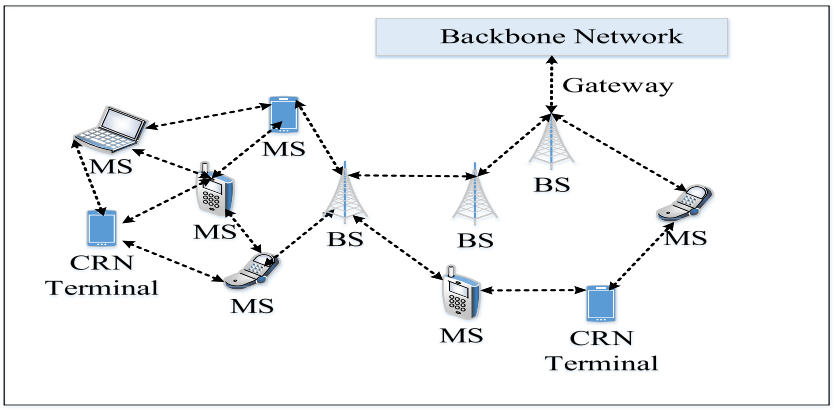


Figure 2.6: Mesh Cognitive Radio Architecture adapted from (Chen et al., 2008)

2.9 The Application of Cognitive Radio Networks

2.9.1 Mesh Cognitive Radio Networks

Multi-hop and/or wireless mesh networks have newly acquired recognition as an inexpensive solution for internet access. Conventional wireless networks are obstructed by wireless bandwidth and security necessary to meet the highest requirements of current wireless applications. "Pragmatic" spectrum access could be used to mitigate problems of insufficient bandwidth in wireless mesh networks, by authorizing mesh nodes to be flexible enough to discover any obtainable spectrum opportunities. Cognitive mesh networks are often used to supply broadband access to regions, sub regions and other areas that suffer from lack of resources (Steenkiste et al., 2009). Figure 2.7 shows the topology of mesh CR network.

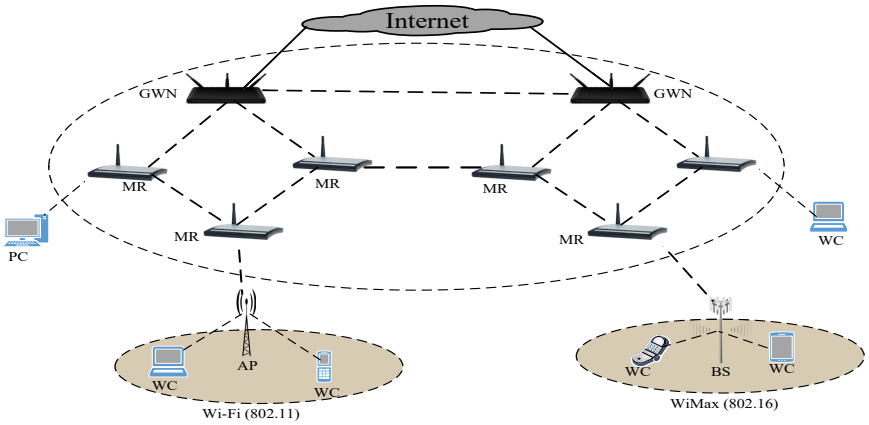


Figure 2.7: Mesh Cognitive Radio Network

2.9.2 Public Safety Networks

Public Safety Networks (PSNs) is another application of Cognitive Radio Networks. Public safety networks are wireless communication networks utilized for disaster and catastrophe relief or by providing assistance to public crises and any other activities requiring rapid and trusted communication. Utilization of PSNs is usually for communication between emergency services personnel, when there is little or no technology to perform vital communications across various spectrum usages. Public safety licenses have a wide diversity of bands obtainable including VHF- low, VHF-Hi, 220MHz.

Cognitive Radio Networks provide public safety networks with bandwidth by using opportunistic spectrum access. Moreover, Cognitive Radio Networks provide fundamental communication enhancement,

by allowing access to various public safety services, by adapting quality and medium quality traffic to utilize the networks in a reliable manner. Public safety networks depend on their goals, status and capacity for mission criticality.

Those networks must assist the interlinking of item such as laptops, hand-held devices, and mobile video cameras. Moreover, they provide easier communication, cooperation, and processing with central leadership, co-workers, and another agencies as well as supporting existing conditions to the fullest extent of the required processing (Steenkiste et al., 2009; McGee et al., 2012). Figure 2.8 depicts how Public Safety Networks (PSNs) operate. Different levels of government need to exchange information and communications when confronted with public safety occurrences. Inter-office cooperation of this nature led to the establishment of PSNs.

2.9.3 Catastrophe Relief and Emergency Networks

A Cognitive Radio Network can be used in disaster management especially those caused by natural disasters such as wildfires, earthquakes, volcanic eruptions, that damage communication infrastructures and disrupt communication services. Hence, they are an imperative for communications during rescues, for planning, coordination and providing relief.

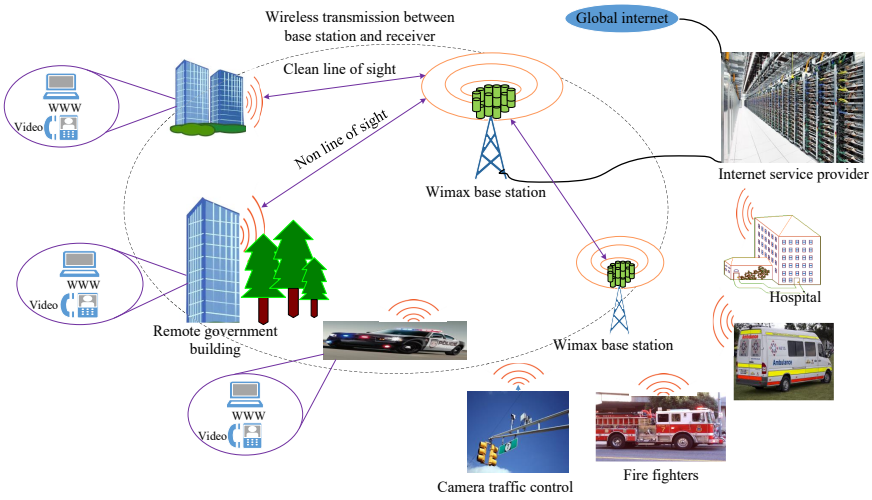


Figure 2.8: Public Safety Network

The existence of a Cognitive Radio Network can assist by utilizing opportunistic spectrum access that provides the required amount of bandwidth and which can support the largely predictable amount of voice, video, critical situation data and traffic sensitive bandwidth effectively and promptly. For example, a Wireless Local Area Network (WLAN) was used for relief communications during the Haiti earthquake. While communication through such networks is often untested and suffers major delays, Cognitive Radio Networks can supply an important bandwidth for voice, and time sensitive traffic (Yücek and Arslan, 2009). Figure 2.9 shows how an emergency network works. It includes Internet access by satellite in areas not accessible to Digital Subscriber Lines (DSL), new architectures for near-on-demand video, satellite networks combined with WiFi, WiMAX, or satellite networks, working together to establish a combined mobile network for Ambulance, trains, and ships (Oliveira et al., 2011).

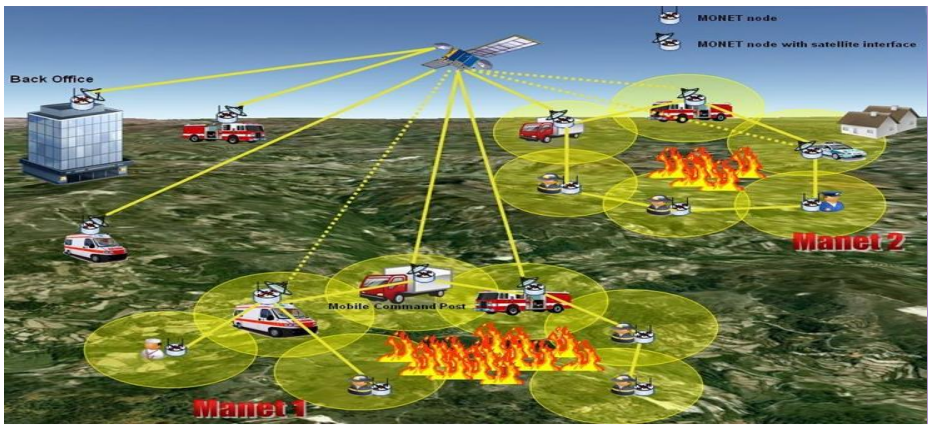


Figure 2.9: Catastrophe Relief and Emergency Network , adapted from (Oliveira et al., 2011)

2.9.4 Battleground Military Networks

The advancement of wireless technology in recent times has made networks easier to hack and for communication signals to be jammed. Therefore, secure communication in battlefields becomes even more challenging for accomplishing a mission. Battlefield network interfaces often provide the only means of communication between soldiers, armed vehicles and other combat units in the battlefields. It becomes the only source of modern communications between soldiers and their commanders. A battlefield network, not only requires a large amount of bandwidth, but also requires secure and reliable communications to transfer dynamic, mission critical information.

CR is one of the principal technologies that enables heavily deployed networks using distributed spectrum access strategies to meet the

bandwidth and reliability requirements. The dynamic nature of Opportunistic Spectrum Access (OSA), a known feature of CR, keeps track of jamming which makes communication on the battlefield difficult. As a result, (DARPA), launched the Wireless Network After Next (WNAN) aimed to create a flexible military communication infrastructure (Khattab et al., 2013). Figure 2.10 illustrates a WNAN battlefield network.

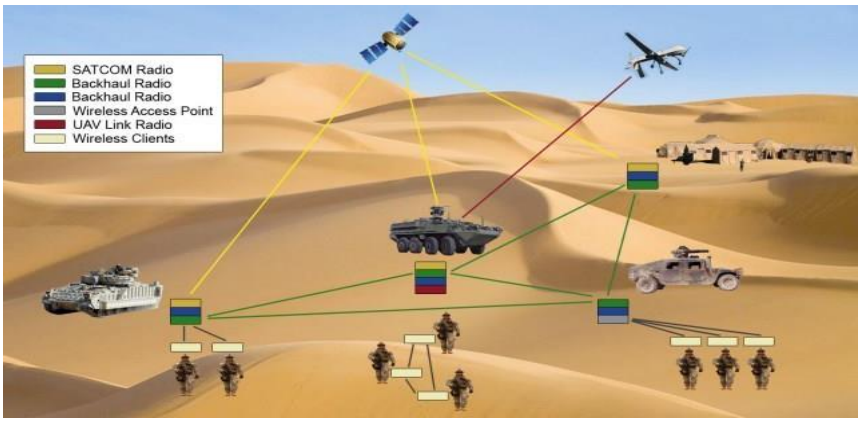


Figure 2.10: Battleground Military Network

2.9.5 Leased Network

All of the applications of Cognitive Radio Networks have secondary users using primary network resources without contributing in any way. A primary network leases a portion of the licensed spectrum for its use, whereas Secondary Users (SUs) use Cognitive Radio techniques to opportunistically obtain use of licensed spectrum. Entry of secondary users to the primary network will lead to raising the cost to primary users of the licensed spectrum (Guijarro et al., 2011). Figure 2.11 illustrates this condition.

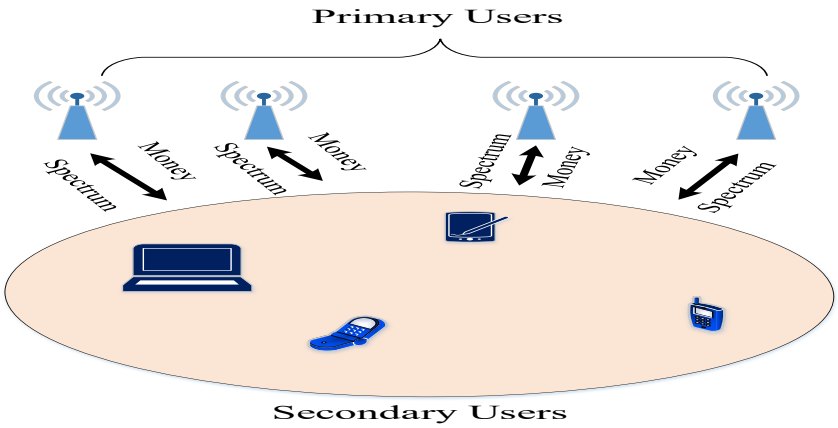


Figure 2.11: Leased Cognitive Radio Networks

2.10 Security of Cognitive Radio

Distributed Cognitive Radio Networks increase threats and other security attentions. CR networks tend to tolerate incipient wireless network weaknesses and threat points but some modern threats can degrade their functions (Idoudi et al., 2014).

Because of its deployment, vital features of Cognitive Radio Networks, and Cognitive Radio, become more vulnerable to attacks and threats. Traditional threats include: eavesdropping, tampering, spoofing and new threats that contain Primary User Emulation (PUE), spectrum managers attacks, and general jamming. Robust security is necessary to create Cognitive Radio networks that are both usable and trustworthy. Countermeasures are necessary to guarantee that secondary and primary users of the spectrum are mostly protected. Threats in Cog-

nitive Radio Networks may be grouped into two classes: traditional and specific threats (Fragkiadakis et al.; Pei et al.; Li and Cadeau; Mao and Zhu; Mody et al.; Safdar and Neill).

The term “jamming” is used to describe the deliberate use of radio noise or signals in an attempt to disrupt communications (or prevent “listening” to broadcasts). The term "interference" is used to describe unintentional forms of disruption versus jamming which is intentional. Jamming is usually directed at radio signals primarily to obstruct. A transmitter tuned to the same frequency as an adversaries’ receiving equipment, and with the same kind of modulation, can, with sufficient power, blanket any signal at the receiver, effectively destroying communication.

The most popular kinds of “signal jamming include random noise, random pulse, stepped tones, warbler, randomly keyed modulated continuous wave, tone, rotary, pulse, spark, recorded sounds, gulls, and “sweep-through”. The primary goal is to bar the correct reception of transmitted signals which causes damage to the receiving operator. Attackers target the physical, data link, network and transport layers but most generally attack in CRN in the physical layer (of the 7-layer OSI communication model).

2.10.1 Traditional Threats

The security of Cognitive Radio networks is one of the most important issues within CRNs. CRNs suffer from numerous security threats because of significant factors in the status of the whole of wireless networks communications. CRNs are vulnerable to security threats produced by communication from wired networks. Threats and attacks on wireless network nodes may include strategies involving eavesdropping, impersonation and traffic analysis. In general, these threats will cause damage to wireless networks, in particular CRNs. Basic attacks and traditional threats include:

- **Eavesdropping Attack**: The attacker monitors network communications to gain enough information about sessions between communicating parties, be they Primary Users, or Secondary Users, and utilizes that information to initiate attacks.
- **Impersonation Attack**: The attackers employ a legal Cognitive Radio node identity, in the wireless network and communicate with another node using this identity. In this situation, the base station provides details of radio nodes in the network and the existence of primary users, not realizing it is dealing with a false user (attacker).

- **Selective Forwarding Attack**: The CR specifies a time limit for data delivery between two radio nodes. If this is exceeded and a PU or SU did not receive the data message, it will inform the BS via another secure radio node. The BS resends the message using the same path. Messages should be encrypted so that any malicious attacker does not obtain valuable data from "lost messages" (Idoudi et al., 2014).

- **Sinkhole Attack**: An insider attack where an invader takes control of a radio node in the network and attacks the network. The node then attempts to attract traffic from neighboring nodes using a routing metric used by routing protocol, which may include contaminated nodes. When able to do so, the contaminated radio node commences an attack. Because of the design of a wireless network, including many to one communications where each node transmits data via the base station, it makes the this Wireless Network (WN) vulnerable to this type of attack (Akan et al., 2009; Kibirige and Sanga, 2015).

- **Wormhole Attack**: The Base Station (BS) provides each radio node with the identities of adjoining nodes and the distances from each of these nodes. All this data should be encrypted. A wormhole attack strives to convince two individual radio nodes

that they are neighbors. This will usually fail when the BS checks its list of IDs and distances, to confirm the action (Idoudi et al., 2014).

- **Hello Flood Attack**: This attack transmits "hello" messages from falsified to legitimate radio nodes, so as to establish links for the provision of false high-quality routes. Some routing protocols utilize a 7-layer OSI communication model and link layer acknowledgments which assists attackers to "spoof" acknowledgments to "push" legitimate radio nodes to debilitated links, which are a subset of passive inactive radio nodes. As an outcome, the links may be accurate for routing but may force packets to transmit to other radio nodes, that may be missing, damaged or malicious (Idoudi et al., 2014).

2.10.2 New Types of Threats in Cognitive Radio

Cognitive Radio Networks suffer from different kinds of attacks that menace primary targets, because of certain operational tasks (Bhattacharjee et al., 2013; Fragkiadakis et al., 2013; Mangai et al., 2013). Some of those attacks include:

- **Software Attacks**: Software attacks have more lasting effects. Software attacks can fully cripple Cognitive Radio networks. Manipulation impedance and virus discovery mechanisms have to be combined to prevent malicious software downloads from trusted servers, as well as direct software attacks. There should be mechanisms to authenticate, authorize, and maintain the integrity of a software installation to prevent attacks.

- **Hardware Attacks**: Tries to destroy the hardware in Cognitive Radio networks or to change their tasks. The effects of the attacks range from closing down a Cognitive Radio completely, to transferring signals to incorrect frequency bands. Moreover they may cause radio nodes to be uninvolved in dynamic spectrum management, cooperation, making decision-making operations difficult, if not impossible. These can lead to insufficient or completely wrong decisions that will effectively compromise the network.

- **Spectrum Sensing Data Attacks**: This provides false spectrum sensing data, which may cause spectral analysis to be inaccurate, leading to incorrect decisions and provide incorrect frequency bands to primary and secondary users. Incorrect bands may expose Cognitive Radio networks activities to damage. If no measures are instigated against these attacks, the transportation

characteristics of different bands may be wrongly determined, and the network may be exposed to more attacks - thus decreasing effective Cognitive Radio network communication.

- **Secondary Spectrum Data Falsification (Byzantine Failure)**: This occurs when radio nodes are incapable of discovering the existence of primary users, because of wrong spectrum sensing data, which occurs as the result of an attack. The main goals of Byzantine attackers is to reduce detection probability, for disrupting normal operation of Primary Users, and raising the probability of false warnings for the purposes of denying access opportunities for correct Secondary Users (Zhang et al., 2015). Secondary spectrum sensing is one of the main methods for enabling Cognitive Radios to learn from their environment and to determine when and where "spectrum holes" are located (Magdalene and Thulasimani, 2017). There are three types of spectrum sensing data falsification attacks:

    1. **Secondary Spectrum Data Falsification (SSDF)**: This attack blocks the application of CR techniques for commerce and military to a very large extent by falsifying secondary spectrum data to SUs.

2. **Induced Secondary Spectrum Data Falsification (ISSDF)**: The attack wrongly announces that a particular communication channel is free, causing interruption to its users and transmitted data to be lost/damaged or compromised.

3. **Sybil based Secondary Spectrum Data Falsification (SBSSDF)**: Attackers in other nodes provide the impression that the legitimate CR radio nodes have preformed sense functions. This causes legal nodes to communicate their data to attacker nodes, by assuming the attackers nodes are legitimately responsible for sensing and communicating correct data about active Primary Users (Idoudi et al., 2014).

- **Jamming Disruption Attack**: Jammers send, often high powered, signals to the BS of the CR, at the same frequency as an authorized transmitter, thus scrambling reception. Effectively, legitimate data transfer is impaired. This is a direct DoS attack for PUs and SUs by interfering with their legitimate communications.

2.10.3 Layers Attacks on Cognitive Radio Networks

There are a number of opinions as to how a Cognitive Radio Network can be secured. One, is that a Cognitive Radio Network can be secured through each layer of the 7-layer OSI communication model, due to layer based attack methods. Attacks on the physical layer are known as a physical layer attack; attacks on the link layer are called the MAC layer (link layer, layer 2) attacks, attacks on the network layer are known as network layer attacks, and attacks on the transport layer are referred to as transport layer attacks (Nanthini et al., 2014; Parvin and Hussain, 2011).

Another method of securing the spectrum channels is by utilizing spread spectrum modulation. Using this approach, the individual characteristics of Cognitive Radio present a protected and harmonious communication (Nanthini et al., 2014; Parvin and Hussain, 2011).

Another view is the use of a digital signature to secure the CRN. An active Primary User Identification (PUI) created from a public key cryptography, is utilized to secure communication through the various channels. (Parvin and Hussain, 2011).

Authentication and encryption methods are used to secure higher layers and the physical layer - on which the CR is strongly dependent, particularly spectrum sensing, making the CR more vulnerable to at-

tacks. Primary User Emulating (PUE) is a method of securing the physical layer utilizing unique information about a communication, commonly called its "fingerprint". PUE fingerprints enhance the security of data transmitted over a multi-path environment which is under physical layer attacks. Wavelet transformation detects and decodes these fingerprints (Nanthini et al., 2014).

2.10.3.1 Physical Layer Attacks

One method of physical layer attack is Primary User Emulation (PUE). The PUE attack occurs by emulating a PU to obtain the resources of specific channels. There is the SELFISH PUE: the aim of which is to amass spectrum resources. This attack is usually executed by two attackers which builds a communication link between the PUE nodes. The PUE attackers attempt to prevent legitimate secondary users from using available frequency bands in the spectrum. If the attack succeeds, the secondary users are effectively blocked from using the spectrum believing that there are little no channels available. The attack disappears when the attackers are either "ejected" or actually leave the network.

CR learns by acquiring information on active PUs and collecting previous behaviours to determine when any channel becomes idle.There are many methods to counter PUE attacks. One is to concentrate on

cross-layer pattern identification. This technique utilities radio signatures of any Cognitive Radio using the spectrum. Waveform identification is utilized to detect malevolent devices. The process includes "the enrollment in gathering data and trial to identify the user" (Nanthini et al., 2014; Chen and Park, 2006; Anand et al., 2008; Pei et al., 2010). This approach is cross-layer security and is able to highlight that characteristic among Cognitive Radio devices. It is also defined as one of the best methods to protect against PUE attacks (Nanthini et al., 2014; Chen and Park, 2006; Anand et al., 2008; Pei et al., 2010).

An objective Function Attack or OFA is by exploiting the radio parameters: "bandwidth, center frequency, modulation type, power, encryption type, protocol, coding rate, frame size, and channel access". Cognitive Radios compute all radio parameters and as a result of some of these parameters may increase or decrease the power for data transmission. Attackers tend to attack during the process of computation. The attackers take control of the computation and biases the outcomes which are tailored to the attackers advantage. Whenever the Cognitive Radio attempts to utilize a higher security level, the attackers jam the spectrum, thus decreasing the CR's overall objective. The result is that the CR will avoid increasing the security level so as not to reduce its objective function (OF) (Anand et al., 2008; Chen and

Park, 2006).

By jamming, the attacker sends interference packets in a continuous stream which impedes the communications of the legal participants. This causes the legal user to either sense the channel as being continuously busy or forces them to receive incorrect or damaged packets. It effectively disrupts network communications for all legal users (Nanthini et al., 2014).

There are two methods of anti-jamming. The first, is to avoid denial of service (DoS) by "channel surfing" or frequency hopping. The second is a "locative escape", where a legal user changes its location to avoid the frequency bands used by the attacker for interference. By this approach, the primary point is to move from the attacker's region to a safer one, where the users must wait until they are within the range of each other to resume communication (Xu et al., 2005; Sampath et al., 2007).

2.10.3.2 Link Layer Attacks

Link layer attacks include:

- **Spectrum Sensing Data Falsification (SSDF)**: This occurs when an attacker transmits incorrect spectrum sensing information to neighboring nodes or to BS. The attacker compels the

receiver to accept incorrect spectrum-sensing decisions. SSDF affects both CR centralized and distributed networks. In a centralized CRN, a head radio node (centre) is accountable for maintaining all sensed network data and decides which frequency bands are busy and which are free. Attacking the centre and falsifying the maintained data, may seriously hinder legal users. This type of attack may be countered by calculating a threshold value from the sum of legitimate sensed spectrum data. If subsequent calculated sums are unacceptably greater than the threshold value, then the channel is unacceptably busy. This means determining an available frequency band (channel) that is most likely to be free from the attacker. If the process continues, coping is by raising the threshold value (Wang et al., 2009).

- **Selfish Channel Negotiation (SCN)**: In a multi-hopping channel system (network), a SCN can force any CR to deny forwarding of data to another network. This attack decreases productivity of the entire Cognitive Radio system. The consecutive likelihood test can be used for this purpose in order to prove its performance in terms of detection time (Zhu and Zhou, 2008; Bian and Park, 2006)

- **Control Channel Saturation Dos Attack (CCSD).**

2.10.3.3 Network Layer Attacks

Network layer attacks include:

- **A Sinkholes Attack**: This presents as the ideal path to a particular destination, attracting neighbour nodes to use it for forwarding their data. An attacker could use this in a way to deliver another attack called selective forwarding, which changes or dismisses data from any radio node in the network. This attack focuses on the infrastructure of a CR and mesh networks. This attack is defeated by geographic routing protocols. The geographic routing protocol attempts to establish a communication topology utilizing only local connections, avoiding the base station. (Hu et al., 2003).

- **A Hello Flood Attack**: The attack is carried out when an attacker transmits broadcast messages to every node in a network containing sufficient data to convince them that it is a neighbour node. When the attack is detected, there is a chance of data loss and non-attendance of legitimate neighbours to forward the data packets. To avoid this attack, each data packet includes a key called a symmetric key, which is checked by a participating trusted base station. The Kerberos algorithm is used in cryptog-

raphy to facilitate the creation of these keys for communication between different radio nodes in the network. To stop an attacker from using an existing session key is to restrict the use of shared keys. The symmetric key is recommended because they are quicker and have a lower overhead on the system (Hu et al., 2003; Wang et al., 2010b).

2.10.3.4 Transport Layer Attacks

Lion Attack: The Primary User Emulation attack is used to disable a Transmission Control Protocol (TCP) connection. It is a cross-layer attack directed at the transport layer (7-layer OSI communication model) where the attacker imitates a licensed transmission sought to achieve a frequency hand-off thus reducing TCP performance. The attacker prevents message transmissions resulting in network starvation (Hernandez-Serrano et al., 2011).

2.10.4 Related Work and History of Multi-Armed-Bandit Problem

Much research has been conducted on the Multi Armed Bandit (MAB) problem. A brief overview:

(Thompson, 1933) conducted a stochastic Multi-Armed Bandit simu-

lation introducing Thompson sampling as an optimal heuristic, which is still an action selection strategy that is superior to more later strategies.

(Robbins, 1985) introduced the initial series analysis of the single player MAB problem. (Bellman, 1956) formulated the Multi Armed Bandit (MAB) as a variant of the Markov decision process (MDP). (Gittins, 1979) demonstrated a Bayesian optimal indexing scheme for the MAB problem, creating a steady MDB. (Lai and Robbins, 1985) presented the concept of "regret", obtaining its lower bound by using the Kullback-Leibler variance built on closely optimal allocation principles.

(Anantharam et al., 1987) extended Lai & Robbins from the single to multi-player. (Whittle, 1988) presented PSPACE - hard "impatient" MAB and showed that sub-optimal indexing is reasonable. (Rivest and Yin, 1994) suggested the Zheuristic which provided the best experimental performance. (Auer et al., 2002) suggested the Upper Confidence Bound (UCB), a hopeful indexing scheme.

2.10.5 Related Work for Jamming Attack in Cognitive Radio

2.10.5.1 Work in Jamming Attacks

Earlier work on wireless jamming attacks has concentrated on several attack models, detection mechanisms, and normal solutions. (Bellardo

and Savage, 2003) showed the vulnerability of the IEEE 802.11 MAC frame to jamming attacks. Using this, off-the-shelf hardware can be employed to perform a variety of attacks. (Xu et al., 2005) showed four jamming attack models with different levels of intelligence, and suggested mechanisms to detect any attack by measuring signal strength, transporter sensing time and packet transmission ratio. (Xu et al., 2004) show a simple technique to alleviate jamming attacks by hopping between channels and physically staying away from the adversary (in a geographic sense). Current work on jamming attacks mostly concentrates on single channel networks on the principle that an enemy can attack a user by following the user as it hops across channels, gaining information about the channels with its aim to jam them.

2.10.5.2 The Theoretical Participation

Theoretical participation often concentrates on proposing various jamming methods, and analyzing their effects under differing system parameters. Usually, the methods are generated by simulators in open source networks simulation, including the Network Simulator 3, OPNET, and OMNEST (Foundation., 2016; Riverbed Technology, 2016; Simulcraft Inc., 2015).

(Xu et al., 2005) presented diverse types of jamming attacks: constant, deceptive, random, and reactive. They studied the attackers,

influences on wireless networks and suggested different methods of an anti-jamming system, used for the purpose of disclosing the existence of jamming attackers. In countering these attacks, signal strength matchmaking and location information checks show them to be more secure methods.

(Sampath et al., 2007) showed CR jamming multiple channels at the same time, on targeted networks, was possible, by a simulation in Qualnet which showed the jamming effects for various numbers of channels, channel switching delays, and the packet size of the jamming method.

(Amuru and Buehrer, 2014) have taken the simmulation methodology to a more comprehensive level, where the jammer includes information on the environment status of "treatment delay".

2.10.5.3 The Experimental Participation

Experimental participation regarding smart jamming attack and anti-jamming is still rare, due mainly to complicated issues correlated to implementing smart Jamming and anti-jamming attacks, under stringent real-time conditions for signal detection and automatic reconfiguration of parameters.

(Wilhelm et al., 2011) have conducted an implementation of a smart

jamming attack on a Universal Software Radio Peripheral2 Software Defined Radio (USRP2) SDR, and have shown the jamming achievements of various jamming signals: narrow band noise, single-tone, and random modulated signals, in an IEEE 802.15.4 based network.

(Nguyen et al., 2014) have conducted an implementation of real-time, protocol aware, reaction jammer target for high-level speed wireless networks. The implementation was executed on a Universal Software Radio Peripheral (USRP N210) SDR. Two compound algorithms were implemented in order to signal detect on the jamming side including: signal cross-correlation and energy detection. They also conducted research on the vulnerability of IEEE 802.16e networks to interactive jamming attacks.

(Liu and Ning, 2012) proposed a BitTrickle anti-jamming wireless communication scheme that permits communication despite the existence of a broadband and high power reactive jammer, by exploiting the reaction time of the jammer. They developed a prototype of BitTrickle using the USRP platform.

2.10.5.4 The Game-theoretical participation

Obviously, there are conflicts of interest between RF jamming and anti-jamming systems. The major aim of the Jammer is to interrupt the effective communication of any data transmission, whereas the

anti-Jammer tries to ensure that communication occurs. So, Game theory - a good framework for analyzing conflicts between responsible players - is an appropriate tool to analyze jamming/anti-jamming issues. Game theory allows for the determination of optimal and near-optimal strategies for jamming and anti-jamming, and to create learning algorithms which are capable to assemble with these strategies.

“ Most recent contributions to the literature on the application of game theory to intelligent jamming problems consider either channel surfing or power allocation as anti-jamming strategies. Furthermore, they are mutually differentiated, mostly by the objective function subjected to optimization (Signal-to-Noise Ratio, Bit Error Rate, Shannon capacity); various forms of uncertainty (user types, physical presence, system parameters); game modulation (zero-sum vs. non-zero-sum, single-shot vs. dynamic) and considered learning algorithms (Q-learning, SARSA, policy iteration)” (Dabcevic, 2015).

(Altman et al., 2007) have assured the presence and singularity of a Nash equilibrium for a group of games and transportation expenses. Moreover, they have acquired an analytical term for the Nash equilibrium and have developed a jamming game such as the “popularization water optimization problem”. They set up 5 channels and analyzed jamming games performed on those channels.

(Wang et al., 2011) have developed issues of jamming attacks with the Primary User (PU) as game, stochastic zero-sum, in which channel switching was regarded as the anti-jamming schema, using a minimax-Q as the learning algorithm. They compared the performance of the formulated constant policy, but with a "short-sighted decision" which did not take into account issues such as environmental dynamics. The algorithm was shown to display reliable performance, in terms of comprehensive spectrum effective channels, at all times.

(Garnaev et al., 2012) and (Buchbinder et al., 2012) have deemed multi-transport power distribution as an effective anti-jamming strategy and also developed zero-sum games. (Buchbinder et al., 2012) have contributed lower bounds on the comprehensive performance of the system for an online learning algorithm. Garnaev et al., 2012 provided verified evidence of the presence and singularity of Nash equilibrium points for a system where deemed players have information that is imperfect on what channel gains have been achieved.

Optimal jamming strategies were studied by (Amuru and Buehrer, 2014) by circulating modification of a jamming waveform, based on the circulate modification, to different types of targeted systems. They showed that the targeted system is one of: Binary Phase Shift Keying (BPSK) or 4-Quaternary Phase-Amplitude Modulation (4-QPAM), which is an optimal jamming signal created by utilizing BPSK. Quaternary Phase-Shift Keying (QP-SK) or 2^4 Quadrature Amplitude

Modulation (16-QAM) which is deployed by a targeted system, and an optimal jamming signal is created by utilizing QPSK.

2.11 Multi-Armed Bandit Strategies

Strategies for the MAB problem, are considered to be solutions to the problem of spectrum sensing access, which is the dilemma of exploration and exploitation of opportunities to mitigate jamming and other attacks. The problem is posed as a gambler having access to a number of slot machines, and determining which machine to play, by how many times this machine was played, for a maximum expected-gain.

$$\text{Number of slot machines} = \text{Number of armed bandits}$$

2.11.1 Upper Confidence Bound (UCB)

The UCB as identified by (Robbins, 1985), has the multi armed bandit problem as a trade-off between exploration and exploitation, where there are a number of experiments with different options. For selected option, a reward is disclosed; for another option, the reward is not disclosed. The goal of the upper confidence bound (UCB) algorithm is to increase the overall reward but reduce the “regret”. (Auer et al., 2002)

proposed a new approach providing a simpler solution for the MAB problem. It includes calculating the top of the upper confidence bound indicator which has gained the attention of researchers in automated learning.

2.11.2 KL-Confidence Bound (KLUCB)

The Kullback-Leibler divergence is used to measure the difference between two probability distributions for the variable X. KL has generally been applied to data mining techniques, but the KL has been used extensively in both probability and information theories. KL divergence is closely linked with entropy in both information divergence and discrimination information. KL is also an asymmetrical measurement between two different probability distributions $p(x)$ and $q(x)$. The KL divergence between $p(x)$ and $q(x)$ indicated by $D_{KL}(p(x), q(x))$, is a measurement of missing information when $q(x)$ is utilized to converge $p(x)$. Both $p(x)$ and $q(x)$ sum up to 1 where, $p(x) > 0$ and $q(x) > 0$ for every x in the distribution X. The Kullback-Leibler divergence for $p(x)$ and $q(x)$ is defined as: (Kullback and Leibler, 1951).

$$D_{KL}(p(x) \parallel q(x) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)} \tag{2.1}$$

The KL divergence measures the predictable number of additional bits

in a data transaction, to digitally code samples from $p(x)$ when utilizing the code dependent on $q(x)$, instead of utilizing code dependent on $p(x)$. $p(x)$ is usually the real distribution of monitored data, or the theoretical one calculated by equation 2.1. $q(x)$ usually explains the theory, pattern, character, or the approximation of $p(x)$.

The Kullback-Leibler divergence is often expressed as follows:

$$D_{KL}(p(x) \parallel q(x)) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx \tag{2.2}$$

KL measures the difference between two distributions. KL is not a measure for distance, because it is not a metrical measure and is asymmetric. The KL of $p(x)$ and $q(x)$ is not commutative. $D_{KL}(P \parallel Q)$ is always positive. $D_{KL}(P \parallel Q) \geq 0$ and $D_{KL}(P \parallel Q) = 0$ when $P = Q$. Note, that interest should be paid when KL divergence, are known the $\lim_{p \to 0} p \log p = 0$. Although, when $p \neq 0$ but $q = 0$, and $D_{KL}(p \parallel q)$ is defined as ∞. That means that there is one possible e event. For example ($p(e) > 0$), and the other expects it to be impossible at all. For example ($q(e) = 0$). Then the distributions are quite different. Nevertheless, from the practice side, p and q distribution are obtained from monitoring sample counting, which is from a frequency distribution. It is unconscionable to expect the obtained probability

distribution, which is the event will be perfectly unattainable since we should take into account the potential of the hidden events. The smooth way can be utilized to pediment the probability distribution from the frequency distribution which was observed.

Example: Assume there are two types of distribution, p and q as follows:

$P : (a : 3/6, b : 1/6, c : 1/6)$ and $Q : (5/7, b : 3/7, d : 1/7)$. To calculate the KL, $D_{KL}(P \parallel Q)$, we offer a small constant ϵ, e.g $\epsilon = 10^{-3}$, and determine smoothing version of P and Q, $P'$and Q' is as follow:

The sample group which is observed in P, $SP = \{a, b, c\}$. Likewise, $SQ = \{a, b, c\}$. Union set will become $SU = \{a, b, c, d\}$. By smoothing, the absent symbols will be added to each distribution according to that, with a small probability ϵ. Then, we have $P' : (a : 3/6 - \epsilon/3, b : 1/6 - \epsilon/3, c : 1/6 - \epsilon/3, d : \epsilon)$ and $Q' : (a : 5/7 - \epsilon/3, b : 3/7 - \epsilon/3, c : \epsilon, d : 1/7 - \epsilon/3)$. $D_{KL}(P', Q')$ (Moreno et al., 2004)

2.11.3 Thompson Sampling

(Thompson, 1933)is the first one who proposed this sampling technique which is used to resolve the multi armed bandit problem. This was later formulated into an algorithm (Toldov et al., 2016). Nev-

ertheless, it was ignored by many researchers applying strategies for the MAB problem, particularly those in artificial intelligence research. Lately, Thompson sampling has been used for various online/communication problems and is also used for resolving jamming attacks on Cognitive Radio networks. (Toldov et al., 2016) defined the Thompson sampling algorithm to solve issues in multi-hop activities in Cognitive Radio networks and found the Thompson sampling strategy a more effective mathematical technique.

3 Theoretical Apparatus

3.1 Game Theory

Game theory is also called match theory. It is defined as a means of mathematical analysis of conflicts of interest, to reach the best possible decision-making options under given conditions, to obtain desired results. In the beginning, game theory was associated with games of chance such as Checks and Poker. But as the theory grew, it was associated with more serious issues/dilemmas in sociology, economics, politics, and military science (Koçkesen and Ok, 2007)

3.2 Multi Armed Bandit

The Multi Armed Bandit problem is an example of problems of successive decisions, with an exchange of exploitation and exploration. An equilibrium of actions which have the highest payoff and explores in-

novated actions is given the highest payoff in the future (Bubeck et al., 2012). The Multi Armed Bandit problem is based on the outcomes of a gambler betting on actions of poker machines.

Study of the Multi Armed Bandit problem dates back to the 1930s,with its exploration, and exploitation trade off being applied to new applications such as advertisement placement, website optimization and packet routing. The Multi Armed Bandit problem defines the payoff operation related to each action.

There are three main forms of the Multi Armed Bandit problem, based on the notion of reward: Stochastic, Adversarial, and Markovian.

A player or forecaster (gambler) compares their performance with that of an optimal strategy, for a horizon of n time steps, they continually play the "best" machine in the first n steps, then study the "repentance" of the predictor for not playing optimally. In details, it is assumed $K \geq 2$ arms, and concatenation $X_{i,1}, X_{i,2}, ...$ of unknown rewards connected with each arm $i = 1, ..., K$ the study predictor at each time step $t = 1, 2, ..,$ choosing an arm I_t and receive the connected reward $X_{I_t,t}$. The regret after n plays $I_1, ..., I_n$ is determined by

$$R_n = \underset{i=1,...,k}{max} \sum_{t=1}^{n} X_{i,t} - \sum_{t=1}^{n} X_{I_{t,t}} \tag{3.1}$$

If a time horizon is not previously known, say, the predictor can at any time, reward $X_{i,t}$ and select I_t - both may be stochastic. This enables the identification of the current regret from either the averaged regret and/or expectation of regret as given in equation 3.2:

$$\mathbb{E}R_n = \mathbb{E}\left[\max_{i=1,\ldots,k} \sum_{t=1}^{n} X_{i,t} - \sum_{t=1}^{n} X_{I_{t,t}}\right] \tag{3.2}$$

Pseudo-regret is determined by:

$$\overline{R}_n = \max_{i=1,..,k} \mathbb{E}\left[\sum_{t=1}^{n} X_{i,t} - \sum_{t=1}^{n} X_{I_t,t}\right] \tag{3.3}$$

The expectation of regret is acceptable for random draw rewards and it is how forecaster's behaviour works. Note: pseudo-regret is the weakest of all regrets. When comparing the anticipated optimal procedure's expected regret, to the regret with respect to the active procedure (optimal on the basis of successive of rewards) the best regret follows the expression: $\overline{R}_n \leq \mathbb{E}R_n$.

The main modulation of (Robbins, 1985) based on the work of (Wald, 1973)and (Arrow et al., 1949) has any arm $i = 1, ..., K$ matched to an unknown likelihood distribution ν_i on $[0, 1]$, and rewards $X_{i,t}$ are

based on draws from distribution ν_i matched to the chosen arm.

3.2.1 Stochastic Bandit Problem

The stochastic bandit model was developed by (Lai and Robbins, 1985). They inserted the method of upper bound for the asymptotic determination of regret. Studies on stochastic bandits, a game theoretic formalization of trade-off between exploration and exploitation, has been investigated individually. In a stochastic bandit problem the focus is to maximize the predicted reward.

Known parameters : number of arms K and probable number of rounds $n \geq K$.

Unknown parameters: K possibility distribution $\nu_1, ..., \nu_K$ on $[0, 1]$. For each round $t = 1, 2, ...$ the predictor selects $I_t \in \{1, ..., K\}$;

Given I_t, the environment draws a reward $X_{I_{t,}t} \sim \nu_{I_t}$ in isolation from the past and provides it to the predictor.

For $i = 1, ..., K$ μ_i indicates the mean of ν_i (the reward of arm i) in the relationships shown below:

$\mu^* = \max_{i=1,...,K}$ and $i^* \in \operatorname{argmax}_{i=1,...,K} \mu_i$.

The pseudo-regret is determined by equation 3.4

$$\overline{R}_n = n\mu^* - \sum_{t=1}^{n} \mathbb{E}\left[\mu_{I_t}\right] \tag{3.4}$$

Example: Assume a fake casino, where per slot machine $i = 1, ..., K$ and stage time $t \geq 1$ the agent adds a reward to $X_{i,t}$, that could be "totalitarian" and perhaps maliciously selected, of value $g_{i,t} \in [0, 1]$. Note it is not in the agent's interest to simply assign all earnings to zero, otherwise gamblers will not use that casino. The forecaster chooses an arm sequence $I_t \in \{1, ..., K\}$at any time step $t = 1, 2, ...$ and monitors the earnings $g_{I_t,t}$. After standard terms, we recall the opponent, or adversary of the mechanism that determines the earning for each arm in sequence. We recall that a player is oblivious if the mechanism is separate from a forecaster's actions. Totally, the opponent might be adapted to the forecaster's former behavior, in which case we are talking about the non-oblivious opponent. For example, the agent in the fake casino may watch the way a gambler plays in order to design even-match vicious sequence of earns. Obviously, the difference between the oblivious and non-oblivious opponent is meaningful only when players are chosen randomly. The opponent can choose the poor sequence of earning, at the beginning of the game by simulating the

future action of the player. However, observe that in the presence of the non-oblivious opponent the interpretation of regret is fuzzy. In fact, in this case set the earning $g_{i,t}$ to arms $i = 1, ..., K$ with the opponent at every step t allowed to rely on past random player action $I_1, ..., I_{t-1}$. We can say in other words, $g_{i,t} = g_{i,t}(I_1, ...I_{t-1})$ for every i and t. We can now compare the regret player's accumulative earning to which he would get via playing the top arm in the first n rounds. However, the player has systematically selected the same arm i in each round, viz $I_t = i$ for $t = 1, ..., n$, the adversarial earns $g_{i,t}(I_1, ..., I_{t-1})$. This may have been different from the one the player has actually faced (Bubeck et al., 2012).

Upper Confidence Bound (UCB) Selection

In UCB suppose the distribution of rewards X meets the requirement that there is a convex function ψ on the originally such that for all $\lambda \geq 0$,

$$\ln \mathbb{E}e^{\lambda(X-\mathbb{E}[X])} \leq \psi(\lambda) \text{and} \ln \mathbb{E}e^{\lambda(\mathbb{E}[X]-X)} \leq \psi(\lambda) \tag{3.5}$$

For instance, when $X \in [0, 1]$ one can pick $\psi(\lambda) = \frac{\lambda^2}{8}$ (3.5). It is known as Hoeffding's lemma. raiding a stochastic multi armed bandit utilizing optimism in face of the principle of suspicion, in order to

do this, will utilize the hypothesis above (3.5) to build the upper bound rate on the average of each arm at some constant confidence scale, then select the arm which looks better below this rate. Here we want a standard idea from convex analysis: the "Legendre-Fenchel transform" of ψ, realized by

$$\psi^*(\varepsilon) = \sup_{\lambda \in \mathbb{R}} (\lambda\varepsilon - \psi(\lambda)) \tag{3.6}$$

For example, if $\psi(x) = e^x$ then $\psi^*(x) = x \ln x - x$ for $x > 0$. If $\psi(x) = \frac{1}{p} \mid x \mid^p$ then $\psi^*(x) = \frac{1}{p} \mid x \mid^q$ for any duo $1 < p$, $q < \infty$ in which $\frac{1}{p} + \frac{1}{q} = 1$

Let say $\hat{\mu}_{i,s}$ be exist the sample average of reward gets by drawing the arm i for s times. Note that since

$\hat{\mu}_{i,s}$ similar to $\frac{1}{s} \sum_{t=1}^{s} X_{i,t}$. utilizing the Markovian inequality, from (3.5) will get that

$$\mathbb{P}\left(\mu_i - \hat{\mu}_{i,s} > \varepsilon\right) \leq e^{-s\psi^*(\varepsilon)} \tag{3.7}$$

Here with a probability at lower 1-δ,

$$\hat{\mu}_{i,s} + (\psi^*)^{-1}\left(\frac{1}{s} \ln \frac{1}{\delta}\right) > \mu_i \tag{3.8}$$

We therefore look at the following strategy, recalled $(\alpha, \psi) - UCB$, whereas $\alpha > 0$ is the input parameter: at time t choose

$$I_t \in \underset{i=1,\ldots,K}{\operatorname{argmax}} \left[\hat{\mu}_i T_i(t-1) + (\psi^*)^{-1} \left(\frac{\alpha \ln t}{T_i(t-1)} \right) \right] \tag{3.9}$$

Here we can show the simple of the bound.

Theorem 1(Pseudo-regret of $(\alpha, \psi) - UCB$). Suppose those reward distribution are acceptable equation (3.5) then $(\alpha, \psi) - UCB$ with $\alpha > 2$ accept

$$\overline{R}_n \leq \sum_{i:\triangle_i > 0} \left(\frac{\alpha \triangle_i}{\psi^*(\triangle_i/2)} \ln n + \frac{\alpha}{\alpha - 2} \right) \tag{3.10}$$

In the status of random variables [0,1], taking into account $\psi(\lambda) = \frac{\lambda^2}{8}$ in equation (3.5) the Hoeffding's Lemma- grand $\psi^*(\varepsilon) = 2\varepsilon^2$, which grant the following pseudo-regret bound

$$\overline{R}_n \leq \sum_{i:\triangle_i > 0} \left(\frac{2\alpha}{\triangle_i} \ln n + \frac{\alpha}{\alpha - 2} \right) \tag{3.11}$$

This special status of variables is bounded we indicate to $(\alpha, \psi) - UCB$ purely $\alpha - UCB$. Evidence, note a first if $I_t = i$, the next three

equations at least one of them should be true.

$$\hat{\mu}_{i^*}, T_{i^*}(t-1) + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{T_{i^*}(t-1)}\right) \leq \mu^* \tag{3.12}$$

$$\hat{\mu}_i, T_i(t-1) > \mu_i + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{T_i(t-1)}\right) \tag{3.13}$$

$$T_i(t-1) < \frac{\alpha \ln n}{\psi^*(\triangle_i/2)} \tag{3.14}$$

In fact, suppose the three equations above are wrong then we get:

$$\hat{\mu}_{i^*}, T_{i^*}(t-1) + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{T_{i^*}(t-1)}\right) > \mu^* \tag{3.15}$$

$$= \mu_i + \triangle_i \tag{3.16}$$

$$\geq \mu_i + 2(\psi^*)^{-1}\left(\frac{\alpha \ln t}{T_i(t-1)}\right) \tag{3.17}$$

$$\geq \hat{\mu}_i, T_i(t-1) + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{T_i(t-1)}\right) \tag{3.18}$$

Which means that $I_t \neq i$. In other words, this leads to

$$u = \left\lceil \frac{\alpha \ln n}{\psi^*(\triangle_i/2)} \right\rceil \tag{3.19}$$

we refer to

$$\begin{aligned}
\mathbb{E}T_i(n) = \mathbb{E}\sum_{t=1}^{n} \mathbb{I}_{I_t=i} &\leq u + \mathbb{E}\sum_{t=u+1}^{n} \mathbb{I}_{I_t=i} \text{ and (3.14)} \\
&\leq u + \mathbb{E}\sum_{t=u+1}^{n} \mathbb{I}\ (3.12) \text{ or } (3.13) \\
&= u + \sum_{t=u+1}^{n} \mathbb{P}((3.12) \text{ is correct}) + \mathbb{P}\ ((3.13) \text{ is correct})
\end{aligned}$$

Therefore, it is enough to bound the probability of events, (3.12) and (3.13) utilizing a union bound and (3.12) to directly gets, $\mathbb{P}$((3.12) is correct)

$$\leq \mathbb{P}\left(\exists s \in \{1, ...t\} : \hat{\mu}_{i^*,s} + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{s}\right) \leq \mu^*\right) \tag{3.20}$$

$$\leq \sum_{s=1}^{t} \mathbb{P}\left(\hat{\mu}_{i^*,s} + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{s}\right) \leq \mu^*\right) \tag{3.21}$$

$$\leq \sum_{s=1}^{t} \frac{1}{t^{\alpha}} = \frac{1}{t^{\alpha-1}} \tag{3.22}$$

Same as the upper bound keep for (3.13). Direct calculations deduced the proof.

Lower Bound

Here we are showing the result of the upper confidence bound, which is basically non-upgradable when the reward distribution is the Bernoulli distribution. For $p, q \in [0, 1]$ where $\mathrm{kl}(p, q)$ is a Kullbaack-Leibler variance between a Bernoulli of parameter p and the Bernoulli of parameter q, defined as following:

$$\mathrm{kl}\ (p, q) = p \ln \frac{p}{q} + (1 - p) \ln \frac{1 - p}{1 - q} \tag{3.23}$$

Theorem 2 (Distribution-dependent lower bound) believe satisfying a strategy $\mathbb{E}T_i(n) = o(n^a)$ for any reward Bernoulli distribution, for any arm i with $\triangle_i > 0$, and any $a > 0$, for any set reward of Bernoulli

distributions the following exists.

$$\liminf_{n\to+\infty}\frac{\overline{R}_n}{\ln n}\geq\sum_{i:\triangle_i>0}\frac{\triangle_i}{\text{kl }(\mu_i,\mu^*)} \tag{3.24}$$

To compare the result with equation (3.11) we are using the following standard variance, following the left side of Pinsker's variance, and on the right side we will see that $\ln x \leq x-1$,

$$2(p-q)^2\leq \text{kl }(p,q)\leq\frac{(p-q)^2}{q(1-q)} \tag{3.25}$$

The proof follows three steps considering the case of two arms as follows:

Step one: Suppose arm 1 is the optimal and arm 2 is the sub-optimal, which is $\mu_2<\mu_1<1$. lets say $\varepsilon>0$. Age $x\mapsto \text{kl }(\mu_2,x)$ one is continued can be found by $\mu_2'\in(\mu_1,1)$ such that the.

$$\text{kl }(\mu_2,\mu_2')\leq(1+\varepsilon)\text{kl }(\mu_2,\mu_1) \tag{3.26}$$

$\mathbb{E}',\mathbb{P}'$are used as notations, when they are merged with estimated to the changed bandit wheresoever the parameter of arm 2 is changed by μ_2'. We need to make a comparison and modify the forecaster

behavior on primary bandits. Particularly, we have proved that with a sufficiently large forecaster probability does not differentiate between the two problems. Then, by utilizing the truth that we hold, a good forecaster hypothesis, we know the algorithm does not create many errors on the modified bandit wheresoever the optimal arm is arm 2. In another words, we have the lower bound on the numbers of times active optimal arm. This is the reason we involve the lower bound number of time and the arm 2 is activated in the primary problem. Now we have a slight change in the marking for the rewards indicated by $X_{2,1}, ..., X_{2,n}$ sequence of random variables acquired when pulling arm 2 for n times, $X_{2,s}$ is a reward acquired from the $s-$th pull. For $s \in \{1, ..., n\}$, let

$$\hat{\mathrm{kl}}_s = \sum_{t=1}^{s} \ln \frac{\mu_2 X_{2,t} + (1-\mu_2)(1-X_{2,t})}{\mu_2' X_{2,t}(1-\mu_2')(1-X_{2,t})} \tag{3.27}$$

Note that for the first bandit, the $\hat{\mathrm{kl}}_{T_2(n)}$ is non-re-normalized empirical rate of the $\mathrm{kl}(\mu_2, \mu_2')$ at time n because in this state the process $(X_{2,s})$ is independent and identically distributed from a Bernoulli of the parameter μ_2. Another significant feature is the following: for any happening $\sigma-$algebra created by $X_{2,1}, ..., X_{2,n}$ holds the follow-

ing measurement to change identity

$$\mathbb{P}'(A) = \mathbb{E}\left[\mathbb{I}_A \exp\left(-\hat{\mathrm{kl}}_{T_2(n)}\right)\right] \tag{3.28}$$

So as to connect the behavior of the forecaster in the first and modulated bandits in the event

$$C_n = \left\{T_2(n) < \frac{1-\varepsilon}{\mathrm{kl}\ (\mu_2, \mu_2')} \ln(n) \text{ and } \hat{\mathrm{kl}}_{T_2(n)} \leq \left(1 - \frac{\varepsilon}{2}\right) \ln(n)\right\} \tag{3.29}$$

Step two: $\mathbb{P}(C_n) = o(1)$. via equations (3.28) and (3.29) can have

$$\mathbb{P}'(C_n) = \mathbb{E}\mathbb{I}_{C_n} \exp\left(-\hat{\mathrm{kl}}_{T_2(n)}\right) \geq e^{-(1-\varepsilon/2)\ln(n)} \mathbb{P}(C_n) \tag{3.30}$$

The shorthand

$$f_n = \frac{1-\varepsilon}{\mathrm{kl}\ (\mu_2, \mu_2')} \ln(n) \tag{3.31}$$

Utilizing the equation (3.29) and Markovian inequality, above means,

$$\begin{aligned}\mathbb{P}(C_n) &\leq n^{(1-\varepsilon/2)}\mathbb{P}'(C_n) \leq n^{(1-\varepsilon/2)}\mathbb{P}'(T_2(n) < f_n) \\ &\leq n^{(1-\varepsilon/2)\frac{\mathbb{E}'\left[n-T_2(n)\right]}{n-f_n}}\end{aligned} \tag{3.32}$$

Notice now that the modulated bandit arm 2 is the perfect unique optimal arm. It is therefore assumed that for any bandit, any arm with optimal i and any $a > 0$, the strategy meets $\mathbb{E}T_i(n) = o(n^a)$; this means

$$\mathbb{P}(C_n) \leq n^{(1-\varepsilon/2)\leq n^{(1-\varepsilon/2)\frac{\mathbb{E}'\left[n-T_2(n)\right]}{n-f_n}=o(1)}} \tag{3.33}$$

Step three: $\mathbb{P}(T_2(n) < f_n) = o(1)$.

Given that

$$\begin{aligned}\mathbb{P}(C_n) &\geq \mathbb{P}\left(T_2(n) < f_n \text{ and } \max_{s\leq f_n} \hat{\text{kl}}_s \leq \left(1-\tfrac{\varepsilon}{2}\right)\ln(n)\right) \\ &= \mathbb{P}\left(T_2(n) < f_n \right. \\ &\left. \text{and } \tfrac{\text{kl}\ (\mu_2,\mu_2')}{(1-\varepsilon)\ln(n)} \times \max_{s\leq f_n} \hat{\text{kl}}_s \leq \tfrac{1-\varepsilon/2}{1-\varepsilon}\ \text{kl}\ (\mu_2,\mu_2')\right)\end{aligned} \tag{3.34}$$

Now utilizing the maximum version of the law for huge numbers: for any succession (X_t) of separated true random variables with favorable

average $\mu > 0$,

$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} X_t = \mu \ a.s.$ means $\lim_{n\to\infty} \frac{1}{n} \max_{s=1,....,n} \sum_{t=1}^{s} X_t = \mu \ a.s.$

Ago $kl(\mu_2, \mu_2') > 0$ and $\frac{1-\varepsilon/2}{1-\varepsilon} > 1$, we concluded that

$\lim_{n\to\infty} \mathbb{P}\left(\frac{\mathrm{kl}\ (\mu_2,\mu_2')}{(1-\varepsilon)\ln(n)} \times \max_{s\geq f_n} \hat{\mathrm{kl}}_s \leq \frac{1-\varepsilon/2}{1-\varepsilon} \mathrm{kl}\ (\mu_2, \mu_2') \right) = 1$

Therefore, as a result of the step two and equation (3.29) we obtain

$$\mathbb{P}(T_2(n) < f_n) = o(1) \tag{3.35}$$

Here calling that $f_n = \frac{1-\varepsilon}{kl(\mu_2,\mu_2')} \ln(n)$, and utilizing equation (3.26) we get

$$\mathbb{E}T_2(n) \geq (1 + o(1)) \frac{1-\varepsilon}{1+\varepsilon} \frac{\ln(n)}{kl(\mu_2, \mu_1)} \tag{3.36}$$

which has deduced the proof (Bubeck, 2010)

3.2.2 Adversarial Bandit Problem

Known parameters: number of arms $K \geq 2$ and probable number of rounds $n \geq K$. For each round $t = 1, 2, ...$

The predictor selects $I_t \in \{1, ..., K\}$, probably with the assistance of

foreign randomization.

Altogether, the opponent chooses a gain vector $g_t = (g_{1,t}, ..., g_{K,t}) \in [0,1]^K$, probably with the assistance of foreign randomization.

The predictor receives the reward $g_{I_t,t}$, while the earning of the other arms aren't observed.

In the adversarial, the main target is to get the regret bounds best-probability or an expectation for any potential random strategies utilizing via a forecaster or an adversary, regardless of the adversary. If the non-oblivious opponent exists this is not a simple mission. That is why bounding the pseudo- regret

$$\overline{R}_n = \max_{i=1,...,K} \mathbb{E}\left[\sum_{t=1}^{n} g_{i,t} - \sum_{t=1}^{n} g_{I_t,t}\right] \tag{3.37}$$

Here we note that choosing random opponents is not important because of bound carry adversary, on the other hand it is essential to allow a random forecaster. The adversarial relation of a bandit problem was primarily suggested as the way of playing an unknown game versus the adversary. The game is supposed to be unknown by the player, who also watches the adversary's shifts in each play; (Banos et al., 1968), deemed the problem of a little-known recurring game,

where in every game the player watches only his own reward. This problem was found to be quite the same as that of the adversarial bandit problem with the non-oblivious opponents.

The third basic model of the MAB problem supposes that the reward methods are not independent and identically distribution neither are adversarial. More accurately, the arms are linked with K Markov Processes. Every arm has it own case space. Every time the arm i is selected in case s, the stochastic reward is pulled from probability distribution $\nu_{i,s}$, and the case of the reward process from arm i turns in Markovian style, based on the latent stochastic transmission matrix M_i. Both reward and new case of the player are detected. On the other hand, the case of arms that have not been selected keep remand. Here we might consider K computing projects that allocate consecutive to the unit work resource; in this situation the case of the project that obtains the resource might be changed. Furthermore, it is usually assumed that the basic stochastic transmission M_i are known. Therefore, the optimal strategy can be calculated through dynamic programming, and the problem is primarily of a calculation nature; the seminal work by (Gittins, 1979) prepares the greedy strategy that can be calculated efficiently.

3.2.3 Markov Bandits

The Markov bandit is similar to the case of the Bayesian bandits, those parametric stochastic bandits, wherever the parameters of the reward distribution are supposed to be pulled from known previous, and the regret is calculated by calculating the mean through the pull of parameters from the previous. The Markov case is linked with the choosing of the arm here, accompanied by an update of the back distribution of the rewards for this arm after a new reward note (Meshram et al., 2018).

The Markov bandits is a standard model in the file of operations research and economics. Nevertheless, the mechanisms utilized in their analysis differ significantly from those utilized to analyze the stochastic and adversarial bandits

3.3 Investigation Strategies

The Multi Armed Bandit made up of K arms, $K \geq 2$, K is a set of arms, it takes numbers from 1 to K, any a arm linked with the unknown eventuality distribution over rewards p_a which has a mean μ_a, if you pull any ath arm, we will get a reward r that it is inspected from p_a, but there is agent who has action of T arm pulls. The agent

should pull, the arm in a serial way to maximize the reward next T arm pulls. The agent will get the reward for pulling arm a_i at the t'th step be $r_{a_i,t}$ that is inspected from p_{a_i} , the goal of the agent is to maximize accumulative reward $\sum_{t=1}^{T} r_{a_i,t}$ (Raja, 2016). Conclusion, the arm rewards are stochastic; we focus on maximizing the overall expectation of the reward,

$$\begin{aligned} OverallReward &= = \sum_{t=1}^{T} r_{a_i,t} \\ E[OverallReward] &= E[\sum_{t=1}^{T} r_{a_i,t}] \\ E[OverallReward] &= \sum_{t=1}^{T} E[r_{a_i,t}] \\ E[OverallReward] &= \sum_{t=1}^{T} \mu_{x_t} \end{aligned} \tag{3.38}$$

whereas $x_t \in \{a_1, a_2, ..., a_K\}$for $t = 1, 2, ..., T$ are action values of pulling the arm sequentially which means reward of action value, the arm with optimal reward will be called the optimal arm, let the optimal a^*, an optimal arm, μ^* as the reward, by different way to maximize accumulative reward through minimizing the accumulative

expected regret.

$$\begin{aligned} Regret &= \sum_{t=1}^{T} r_{a^*,t} - r_{a_i,t} \\ E[Regret] &= E[\sum_{t=1}^{T} r_{a^*,t} - r_{a_i,t}] \\ E[Regret] &= \sum_{t=1}^{T} E[r_{a^*,t}] - \sum_{t=1}^{T} E[r_{a_i,t}] \\ &= T\mu^* - \sum_{t=1}^{T} \mu_{x_t} \end{aligned} \tag{3.39}$$

Therefore, the agent can minimize the expected accumulative regret if the optimal arm can be determined. At first the agent knows nothing about the distribution of reward from each arm; they have to explore by pulling the arms in some sequence and to learn these distributions. Also the agent has to exploit the pulling arm, which he thinks is more rewarding, since his goal is to maximize the accumulative reward. But how does an agent choose when to explore and when to exploit. A lot of exploration and the agent will accumulative a lower reward. A lot exploiting and the agent may not detect an optimal arm and still accumulate a lower reward.

Let $Q_a$demonstrate the experimental average of the reward by pulling the incoming arm a. Q_a is the unbiased rated of μ_a

$Q_a = \frac{\text{Some of reward incoming from arm } a}{\text{Number of time arm } a \text{ was pulled}}$

The type of Multi-Arm bandit we see here is a stochastic MAB. In this situation, with the probability distribution p_a is [0,1], so μ_a and the

reward are bounded to each arm between [0,1]. A sub-situation of the stochastic bandit is the Bernoulli bandit, where rewards are either 0 or 1 and the probability distribution p_a is Bernoulli distribution with unknown winning probability μ_a(Raja, 2016).

The Bernoulli MAB algorithm is as follows:

Algorithm 3.1 ***Bernoulli MAB algorithm***

1: **Begin**
2: **for** a in 1.....K
3: $Q[a] = 0$, $N[a] = 0$, $F[0] = 0$
4: **end for**
5: **for** t in 1.....T do
6: $a = SelecteArm(Q, N, S, F)$
7: $r = BernoulliReward(a)$
8: $N[a] = N[a] + 1$
9: $Q[a] = Q[a] + \frac{1}{N[a]}(r - Q[a])$
10: $S[a] = S[a] + r$
11: $F[a] = F[a] + (1 - r)$
12: **end for**
13: **end**

Whereas in $Q[a]$ the average experimental reward to pull the arm a,

$N[a]$ is the number of times the arm is pulled,

$S[a]$ is the number of times a reward of 1 was earned when arm a was pulled, and

$F[a]$ is the number of times a reward of 0 was earned when arm a was pulled.

Here now we have to choose how to select the arm in order to bal-

ance exploration and exploitation so that the accumulative reward is maximized.

3.3.1 Random Selection

In the random selection strategy, we will select the arm quite randomly, this is not a useful strategy because it completely ignores the date of pulling the arm; we are only looking at the pulling of random arms to form a baseline for comparison with another strategies. The accumulative regret expected to select a random arm would be as follows:

$$E[Regret] = T\mu^* - \sum_{t=1}^{T} E[r_{a_i,t}] = T(\mu^* - \bar{\mu}) \tag{3.40}$$

Whereas $\bar{\mu}$ is the average of $\mu_1, \mu_2, ..., \mu_K$. and the regret is linear.

3.3.2 Greedy Selection

The Greedy selection almost widely reaches a sub-optimal solution, and is one of the investigation strategies which makes the best selection of the highest reward at a small stage randomly in order to get optimal

rewards. The equation would be.

$$A = argmax_a(Q[a]) \tag{3.41}$$

The initial value for all $Q[a]$ are 0, the arms are chosen randomly until one of them gives a reward of non zero. In this case only that arm will be selected. This strategy is not worth exploring at all so it is highly far-fetched that the arm will be selected. The expectation of accumulative regret for greedy selection arm would be

$$E[Regret] = T\mu^* - \sum_{t=1}^{T} E[r_{a_i,t}] = T(\mu^* - \mu_{a_{i'}}) \tag{3.42}$$

whereas $a_{i'}$ is the arm, where the arm gives non zero rewards, maybe this arm is the first to give nonzero reward, include this expectation of it. The greedy algorithm linear regret as same random selection, the weaknesses of greedy selection method, constantly exploit the existing knowledge with no exploration, and can then be stuck with the sub-optimal action.

3.3.3 ϵ-Greedy Selection

The probability will be $\epsilon < 0 < 1$, when the arm is selected randomly, otherwise the probability will be 1-ϵ when selecting the greedy arm. So, by selecting the greedy arm the probability would be $1 - \epsilon + \frac{\epsilon}{K}$; this means that if any another arm was selected the probability of it would be$\frac{\epsilon}{K}$, will get the function of probability (in other words if the probability is $1 - \epsilon$, select the action with a maximum value, but if the probability is ϵ, select the action randomly from all actions with equal probability)

$$p(a_i) = \begin{cases} 1 - \epsilon + \frac{\epsilon}{K} & , \quad \text{if} \quad a_i = argmax_a(Q[a]) \\ \frac{\epsilon}{K} & , \quad otherwise \end{cases} \tag{3.43}$$

Suppose ϵ is constant, the expectation of accumulation of regret will be as follows

$$E[Regret] = \sum_{t=1}^{T} E[r_{a^*,t}] - \sum_{t=1}^{T} E[r_{a_i,t}] \tag{3.44}$$

$$E[Regret] = \sum_{t=1}^{T} E[r_{a^*,t}] - \sum_{t=1}^{T} E[r_{a_i,t}] \tag{3.45}$$

$$\geq \sum_{t=1}^{T} \mu^* - [(1-\epsilon)\mu^* + \epsilon\bar{\mu}] \tag{3.46}$$

$$= T\epsilon(\mu^* - \bar{\mu}) \tag{3.47}$$

The ϵ-greedy is linear as a Greedy algorithm. We can alternate over the time until the agent acts greedily within his limit as $T \to \infty$. At first, the agent should make randomly to promote the investigation and with the advance of time, the agent should act extra greedily, it has the potential to realize logarithmic regret employing the strategy of decaying -greedy; however, the decay stream requires enough knowledge of $p_i s$ (Raja, 2016).

Example: assume after your first 10 pulls, you select and played machine number 1 four times and won 1$ twice and 0$ twice, the probability of machine number 1 is $\frac{2\$}{4} = 0.50\$$. And assume you played machine number 2 five times and won 1$ three times and 0$ two times, the probability of machine 2 is $\frac{3\$}{5} = 0.60\$$, and you played machine number 3 three times and you won 1$ once and 0$ twice, the probability of machine 3 is $\frac{1\$}{3} = 0.33\$$, now you can choose the machine to play assume machine number 13, here you create a random

number p, bounded between 0.0 and 1.0, assume you have set $\epsilon = 0.10$, if $p > 0.10$, will choose machine number 2, because it contains the maximum average payout, if $p < 0.10$, you choose randomly the machine. Consequently, any machine has $a\ \frac{1}{3}$ chance to be chosen, you will note that the machine number 2 may be chosen anyway because it was randomly chosen from all machines. With the passage of time, the best machine will be played more often, because will it pays out more; in other words the $\epsilon-$greedy selection is the best option (Greedy) more than once, but selecting the random option with the small probability (ϵ) sometimes (Mccaffrey, 2018).

3.3.4 Boltzmann Exploration

The Boltzmann exploration, in a perfect world likes to exploit all the information and to pick the arm a with the probability in the rated $Q[a]$-value; this leads us to action selection known as Soft-Max action. Boltzmann distribution is used when the probability of picking arm a is commensurate to $\exp(Q[a]/\tau)$, the Boltzmann will picks arm a with probability

$$p(a) = \frac{\exp\ (Q[a]/\tau)}{\sum_a \exp\ (Q[a]/\tau)} \tag{3.48}$$

Whereas $\tau>0$ is the temperature how to select arm randomly, when the temperature is higher τ, the arms are selected approximately in equal quantities and, in the limit as $\tau\to\infty$ the optimal arm is continuously selected. Also the Boltzmann exploration gives linear expected accumulative regret (McFarlane, 2018).

3.3.5 Upper-Confidence-Bound Arm Selection

The strategy of UCB is optimism in the face of uncertainty, as we know $Q[a]$ is equitable rated of μ_a. After a few $N[a]$ pulls of the arm a, we can be genuinely sure of how near $Q[a]$ is to μ_a. Utilizing Hoeffding's contrast, which has ben utilized as a part of my past can infer the accompanying bound.

$$p_r(\mid Q[a] - \mu_a \mid \geq \epsilon) \leq 2\text{exp } (-2N[a]\epsilon^2) \tag{3.49}$$

Utilizing a single-sided variant of this inequality we get.

$$p_r(\mu_a \geq Q[a] + \epsilon) \leq \text{exp } (-2N[a]\epsilon^2) \tag{3.50}$$

So arm a, which average reward $Q[a]$ after pulling it $N[a]$ times,μ_a a

upper confidence bound (UCB) which exceeds probability

$$p = \exp\ (-2N[a]\epsilon^2) \tag{3.51}$$

We need the probability that μ_a exceeds UCB to reduce with t, numbers of arms which are pulled until now. We can utilize $p = t^{-4}$, ensuring that we choose the optimal action in the limit as $t \to \infty$.

$$\epsilon = \sqrt{\frac{-\log p}{2N[a]}} = \sqrt{\frac{2\log t}{N[a]}} \tag{3.52}$$

The UCB strategy is accordingly followed:

$$A = argmax_a \left[Q[a] + \sqrt{\frac{2\log t}{N[a]}} \right] \tag{3.53}$$

The UCB strategy can also be considered as a strategy to explore incentive. Here is the further incentive regardless of expected reward to pull the arm a is $\sqrt{\frac{2\log t}{N[a]}}$ which can be described as a reward to earn confidence by rewarding the arm's a, this incentive my cause the agent to pull non-greedy when he thinks he can get extra information about the arms reward (Raja, 2016; Lattimore and Szepesvári, 2018).

3.3.6 Thompson Sampling Strategy

Thompson sampling strategy works by conserving before on average reward by arms μ_i. It gathers values for each arm of its predecessor and selects the arm at the highest value. When the arm a is pulled and notes a Bernoulli reward r, amend the former on reward basis, this step reiterated for the following arm pull. Distribution of Beta is the suitable selection of priors to obtain the Bernoulli rewards. Probability of density for distribution of Beta function with parameters α and β is (Russo et al., 2017):

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \tag{3.54}$$

The Thompson sampling strategy at first supposes arm a to have a predecessor $Beta(1,1)$ on , which is a regular distribution on (0,1). Distribution of Beta is beneficial for Bernoulli in order to, if the predecessor is a distribution $Beta(\alpha,\beta)$, after monitoring Bernoulli experiment, the back distribution is $Beta(\alpha,\beta)$ if the experiment was the winning or $Beta(\alpha,\beta+1)$, if it was the fail. At time t having monitored $S[a]$ win and $F[a]$ fail out of $N[a]$ pulls of arm a, the Thompson algorithm updates the distribution of beta on a as $Beta(S[a]+1,F[a]+1)$.

Thompson algorithm then samples these background distribution of a $\mu'_a s$, the arm is played according to the probability of the largest (Russo et al., 2017; Raja, 2016). The algorithm is accordingly followed:

For a in 1.....K : $\theta[a] \sim Beta(S[a]+1, F[a]+1)$

$A = argmax_a(\theta[a])$

The accumulative regret is expected for Thompson it is a logarithmic sampling strategy.

Example:best example for Thompson sampling is called $Beta - BernoulliBandit$. Assume we have a arms, when pulled an arm a will get reward of one with probability θ_a and the reward of zero with probability $1 - \theta_a$. Each θ_a could be explicated as a probability of success or reward mean. The rewards mean $\theta = (\theta_1, ..., \theta_K)$ are unknown, but it is constant all the time, in the the first time, the action X_1 is applied, and the reward $r_1 \in \{0, 1\}$ is created with success probability $\mathbb{P}\ (r_1 = 1 \mid X_1, \theta) = \theta_{x_1}$. After monitoring r_1, the agent played other action x_2, monitor the reward r_2, and so on continue the process (Russo et al., 2017):

The algorithm for Beta-Bernoulli Bandit is as follows:

Algorithm 3.2 ***Algorithm Bernoulli Thompson*** (a, α, β)

1: **Begin**
2: **for** $t = 1, 2, ...$**do**
3: # sample model
4: **for** $a = 1, ..., a$ **do**
5: Sample $\hat{\theta}_a \sim$beta (α_a, β_a)
6: **end for**
7: # select and play action
8: $x_t \leftarrow argmax_a \hat{\theta}_a$
9: play x_t and monitor r_t
10: #update distribution :
11: $(\alpha_{x_t}, \beta_{x_t}) \leftarrow (\alpha_{x_t}, \beta_{x_t}) + (r_t, 1 - r_t)$
12: **end for**
13: **end**

Testing the above strategy (random selection, greedy selection, ϵ-greedy selection, Boltzmann selection, UCB selection, and Thompson sampling), utilizing 10 arms randomly Bernoulli Bandit with average arm rewards resulting from distribution sampling between [1,0], evaluate these strategy by drawing the mean percentage of the optimal arm pull versus the number of pulls, the average was taken of randomly created cases (Raja, 2016). The random selection pulls an optimal arm only 10% pulls. Greedy selection proves the optimal arm only 20% of pulls. ϵ-Greedy selection is fastest to fine the optimal arm so only pulls 60% of time, The UCB strategy was slow to find the optimal arm but ultimately outperforms the ϵ-Greedy. and the optimal arm almost 100% of the time was pulled by Thompson sampling which is so far, the better strategy. Figure 3.1 shows the result.

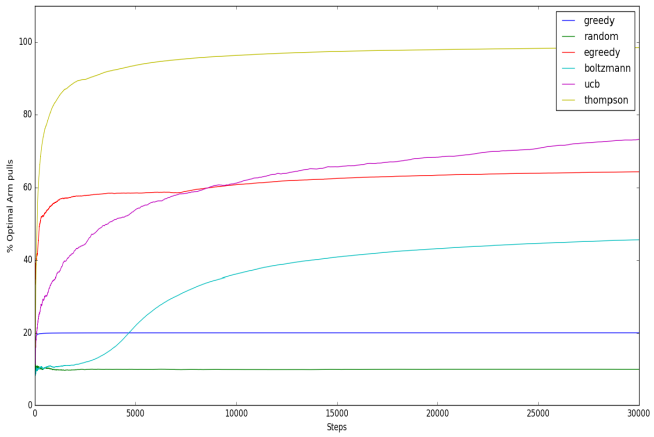


Figure 3.1: Testing Resulting of Strategies (Raja, 2016)

4 Methodology

4.1 Introduction

This section includes the determination of a methodology, designed to ensure continuous communication accessibility between transceivers in Cognitive Radio nodes by mitigating jamming threats. The pivot algorithms used in the methodology include the: Multi-Armed Bandit strategies, Upper Confidence Bound (UCB), Kullback-Leibler Upper Confidence Bound (KLUCB) and Thompson sampling. Also included is a machine learning algorithm. The MAB strategies are considered as a solution to the issue of spectrum sensing access, which solves the problem of exploration and exploitation, also the problem stated when a gambler is given a number of slot machines, and attempts to locate the optimal machine to play and how many times to play the machine.

4.2 The Communication model in Cognitive Radio

Figure 3.1 illustrates a time slotted, multi-channeled opportunistic spectrum open access. There are N channels not overlapping which are

located in center frequency F_e with bandwidth B_e for e = 1,2,.....,K and time T, the time-frequency slot. They are presented by F_e, B_e, T, they give transmission (Ty) opportunity and have a duration of t, supposing the node can sense a neighbor node or other nodes broadcasting in range, like sensing ability, while it is not linked to specific traditional media access control mechanisms.

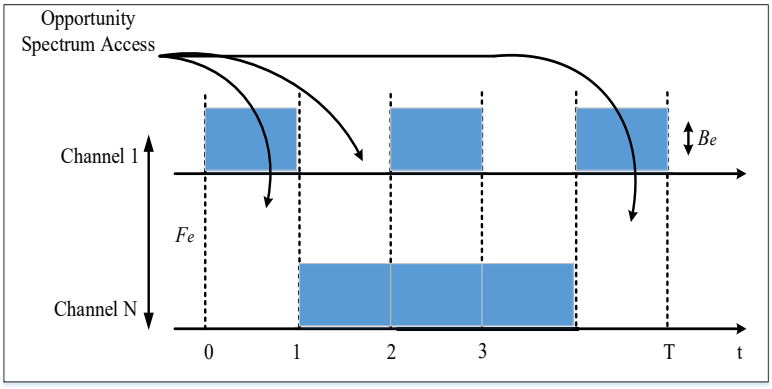


Figure 4.1: Transmission of Opportunity in Slotted Multi-channels Spectrum

4.3 The Multi-Armed Bandit Model in Cognitive Radio

Figure 3.2 describes the MAB algorithm for the Cognitive Radio with the jamming threat in both centralized and distributed networks. The arms are compatible with channels in the spectrum. The communication nodes and jammers both "play" in the network (e.g communication jammers). Networks are multi-nodes. The problem is classified as

a multi-players MAB, which is a variant of the classical single-player MAB (Lai and Robbins, 1985; Anantharam et al., 1987).

There are system differences according to whether the system is a centralized control network or a distribution network.

In a centralized network control decisions are made based on information from the base station and the nodes of the network and distribution of all of node activities. MAB, the decision maker, collects data (on transmissions) from each player (node) and the final results of each play determines the decisions made.

In a distributed system the decision makers are the nodes. Every node makes its own decision depending on data collected based on its best attempt at transmission. In a distributed system, communication is made on a narrow inner network which is used to collect the data and distribute decisions (strategy). For each single MAB play, the player (node) analyzes collected data, calculates rewards, and preserves the play statistics (decisions) which are shared with others in the network (Gwon et al., 2013).

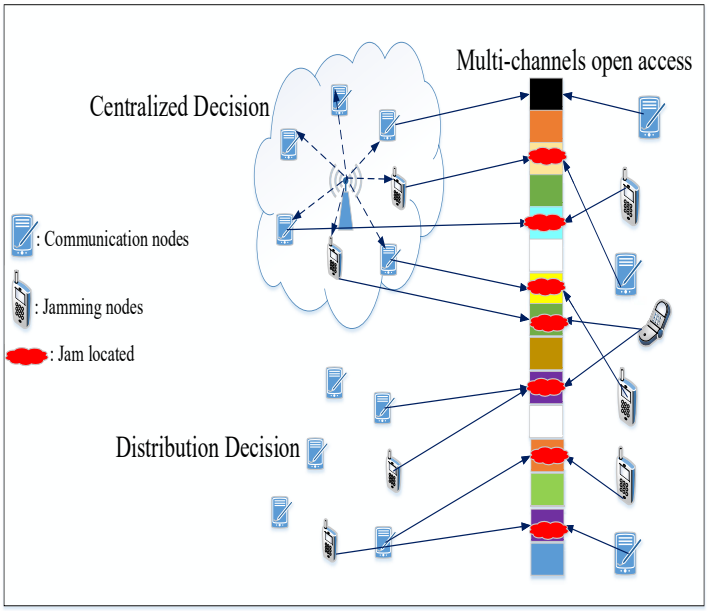


Figure 4.2: Centralized and Distribution Multi-Player Multi Armed Bandit, adapted from (Gwon et al., 2013)

4.4 Proposed method

In this section, the proposed method is explained. The method is constructed from Multi Armed Bandit strategies, confined to classes from the Upper confidence bound, KL upper confidence bound, and Thompson sampling strategies.

4.4.1 Multi-Armed Bandit

The stochastic multi-armed bandit problem represents a significant model to study exploration-exploitation trade-offs, enhance learning for the MAB algorithms - which are well known and understood theo-

retically and have experimental verification that their efficiency exists.

(Robbins, 1985) introduced the Multi-Armed Bandit problem. From that time strategies originating from the Multi-Armed Bandit have been widely used to design models, trading-off by obtaining new information from its surroundings which exploits this to form a consistent and reliable model of the environment it purports to represent. MAB strategies offer natural theoretical modulation in analyzing trade-offs for exploration and exploitation in CR networks. (Berry and Fristedt, 1985) provided a general summary of the MAB problem from a statistical perspective.

4.5 Multi Armed Bandit Policies

MAB strategies do not include state transitions. Therefore, reward is based on action $a(r)$. As a result the MAB learning agent must explore the entire set of actions. In other words, the MAB is stateless. every arm has a stable distribution of reward and it does not rely on actions of arms that have been "played" before. The target of the MAB is to explore the distribution of rewards for all arms and only continue playing the best.

The Markov Decision Process includes the "probability of transition status" $p(x_{t+1} \mid x_t, a_t)$. In addition, rewards follow distribution states

dependent on state transition actions $r(x_t, a_t, x_{t+1})$.

An example of the difference between MAB strategies and the Markov Decision Process, can be best illustrated by the following scenario:

A researcher has discovered a new drug to treat Alzheimer's disease. He believes the new drug is better than other drugs on the market. A doctor can prescribe any authenticated drug to his patients and measures the drugs performance by how effective they are. In a clinical trial, a doctor prescribes a drug to a variety of his patients, rather than just to one of his patients, to maximize the outcome of the performance of the drug on his patients. In applying the difference between MAB strategies and the Markov Decision Process:

- Two drugs=Two arms
- Patients = Experiments
- Drugs performance =Reward

This scenario utilizes a MAB problem and provides the policy for which treatment to select for every experiment. If the doctor selects a sub-optimal drug, its performance will be less than another in treating a patient and the trial's performance will be reduced.

A second example is in the following: a city is selected and all significant places in the city are identified by numbers as per figure 5.1.

A visitor to the city knows nothing about the city's layout (has no map). The visitor is at place 13 and wishes to go to place 21, which can be accomplished in a variety of ways. If places in the city are considered to be nodes in a net, with the rule that there are no diagonal paths, then from place 13 there is only a path to places 12 or 14. But from place 1 there are 4 paths: North, East, West and South. If driving, a vehicle requires X amount of fuel to go from one node to another, within the city. For example, in moving from place 13 to 12, X = 0.05 litres. The problem now is to move from place 13 to 21 following a route that minimizes fuel consumption. For this example, MAB strategies and the Markov Decision Process are employed for determining the strategy:

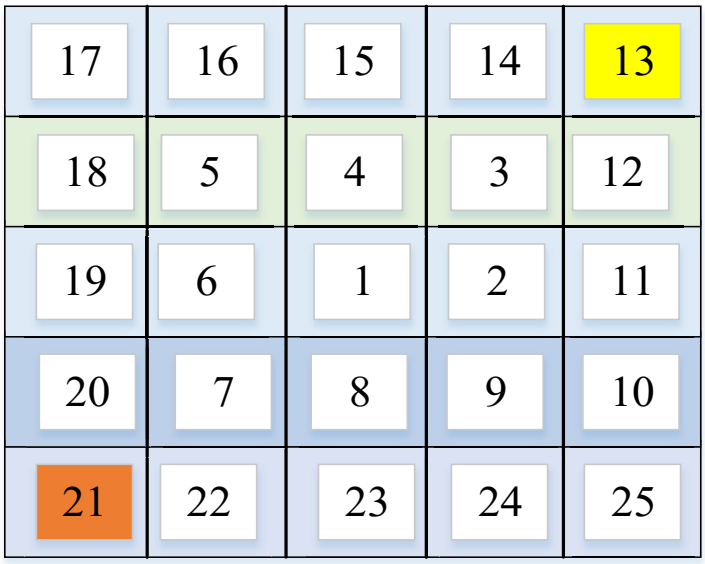


Figure 4.3: Different Ways to Minimize Fuel Consumption

- Place =State
- Action =The direction which could be selected from every place. That means it relies on a situation (place), all the actions vary (e.g, if in place 13, there are 2 actions (paths to 12 or 14). But, at place 4, there 4 actions (or paths).
- Reward =fuel consumption (less or more depending on path)
- Transition probability = Traffic jams, road accidents, safety blockages. These events could be measured for each state (probability for each state). This can be considered as the probability of transition. Unlike MAB, MDP will be in a variable state for each

and every action. MDP's main aim, is to discover the policy that increases (expected) discounted rewards over a unlimited horizon - in this case it is minimizing fuel consumption (Even-Dar et al., 2002).

4.5.1 Adaption Upper Confidence Bound (UCB)

The Upper confidence bound algorithm (UCB) depends on the balance between optimism and pessimism in decision making. That is, actions are dependent on its environment. For a decision to be credible within a MAB strategy, it must be based on the average of payoffs (often unknown) for a number of previous arm plays whose past results have been collected and analyzed. Groundless optimism will not be successful. When you act optimistically either of two things occurs. When optimism is justified, the player behaves optimally. When optimism is unjustified, the player selects actions on the basis of a large reward, which rarely occurs. If this happens often enough, the player learns what is the optimal payoff of an action. In applying UCB to MAB strategies, the actions $X_1, X_2, ..., X_n$ are autonomous and sub-gaussian, meaning that the expected action $\mathbb{E}[X_i] = 0$ and the independent and identical distribution $\hat{\mu} = \sum_{t=1}^{n} X_t/n$. Thereafter, the probable distribution is $\mathbb{P}(\hat{\mu} \geq \epsilon) \leq \exp(-n\epsilon^2/2)$ (Lai and Robbins,

1985; Agrawal, 1995; Katehakis and Robbins, 1995; Auer et al., 2002).

4.5.2 Adaption KL-UCB (Kullback-Leibler Upper Confidence Bound)

The KL-UCB algorithm does not require any tuning or MAB strategy. This algorithm has many supporters including (Filippi, 2010). The KL-UCB algorithm uses a similar procedure as MDP and its analysis employs Bernoulli's probability distribution principles.

Theorem 1 shows the KL-UCB regret is

$$\limsup_{n\to\infty} \frac{\mathbb{E}\left[R_n\right]}{\log(n)} \leq \sum_{a:\mu_a<\mu_{a^*}} \frac{\mu_{a^*}-\mu_a}{d(\mu_a,\mu_a)} \tag{4.1}$$

where $d(p,q) = p\log(p/q) + (1-q)\log((1-p/(1-q))$ indicates to the KL divergence the two Bernoulli parameters p and q. Then Theorem 2, which is said to be a "non-asymptotic, upper-bound on the number of plays of the sup-optimal arm a: for all $\epsilon > 0$ there occurs $C_1, C_2(\epsilon)$ and $\beta(\epsilon)$ such that"

$$\mathbb{E}\left[N_n(a)\right] \leq \frac{\log(n)}{d(\mu_a,\mu_{a^*})}(1+\epsilon) + C_1\log(\log(n)) + \frac{C_2(\epsilon)}{n^{\beta(\epsilon)}} \tag{4.2}$$

Despite the existence of divergence d, it is not fixed for the Bernoulli distribution but is applied to all probable rewards bounded by $[0,1]$.

Pinsker's inequality $d(\mu_a, \mu_{a^*}) > 2(\mu_a, \mu_{a^*})^2$, provides Kl-UCB with a more preferable, theoretical warranty than UCB (which has the same application scope). This improvement is not necessarily observable in simulations. KL-UCB is a first strategy indicator of the lower bound for binary rewards (Lai and Robbins, 1985) but realizes less than the Upper Confidence Bound Variance (UCB-V) regret for the same situation.

The KL-UCB is a generic procedure for bounded bandits and provides the best solution for a binary condition. Kl-UCB is easily adapted within some bandit frameworks, that are not bounded, when the distribution of rewards is probable (likelihood methods). Once divergence d is changed, an optimal algorithm can be built for a wide range of binary situations.

Lemma 9 illustrates how Bernoulli variables are used by the KL-UCB. Theorem number 10 illustrates how an effective trend is used to build trusted instances of bounded variables outputs. Numeral testing confirms the important characteristic of the KL-UCB procedure when applied to MAB strategies. The method is not only superior to MOSS, UCB-V and even UCB in different scenarios, but is comparable to DMED for Bernoulli distributions - especially those for small or mild horizons (Lai and Robbins, 1985; Garivier and Cappé, 2011).

4.5.3 Adaption Thompson Sampling

Thompson sampling strategy is one of the oldest research methods to solve multi-bandit problems. It is a random strategy based on Bayesian principles. Some recent research studies have shown that it provides better performance than some modern ones. The Thompson strategy uses sampling and probability matching, originating in the 1930s, and is specifically used in experiments to solve two armed bandits problem (Wyatt, 1998; Strens, 2000). The strategy showed strong and good empirical performance in clinical trials (Chapelle and Li, 2011; Scott, 2010). In recent years the Thompson strategy has attracted a large amount of literature and has been successfully applied to many applications including management, marketing, web site optimization and Monte Carlo tree searches.

5 Experimental Work

This chapter shows how this research employed Multi-Armed Bandit strategies for protecting against and avoiding jamming attacks on Cognitive Radio network environments. It also includes experimental material on the use of ProML, a method employed for Cognitive Radio jamming attack and protection by simulation, using machine learning approaches.

All simulations were implemented by using the Python programming language tool. The language is a high level programming language, originated by Guido van Rossum in 1991 and expanded by the Python Software Foundation. It was expanded primarily to include reading of symbols, and permitting programmers to express concepts in a small code line count - hence its alternate description: a scripting language. It is a language that enables fast development and combines differing systems more effectively. This language was also selected because it contains libraries that support networking simulation and is more effective in building programming networks.

5.1 Design of Experiment

Figure 5.1 illustrates the design of the experiment, which includes strategy servers, ontology strategy, radio hardware and software, interfaces and spectrum sensing.

Cognitive Radio Software and Hardware: Receives and sends data from System Strategy (SS), as well as transmitting and receiving data from the network.

Reasoning: This means interfacing with the strategy based platforms (see lowest tier in Figure 5.1). Receives delivery strategies from the SS, translates them within a transmission request and transfers them. Receives the transmission answers from the policy reasoner and reroutes them to the SS.

Strategies Server: The strategy server manages the spectrum ontologies and strategies produced by spectrum regulative entities.

Ontology Based Strategy: Carries out spectrum ontologies and strategy based tasks, including loading and processing spectrum ontologies and strategies. It tests the coherence of spectrum ontologies, and evaluates transmission requests against currently active spectrum strategies, to test conformance with those strategies. The evaluation outcome (reply sent) is transmitted to the Reasoning Interface (Bahrak et al., 2012).

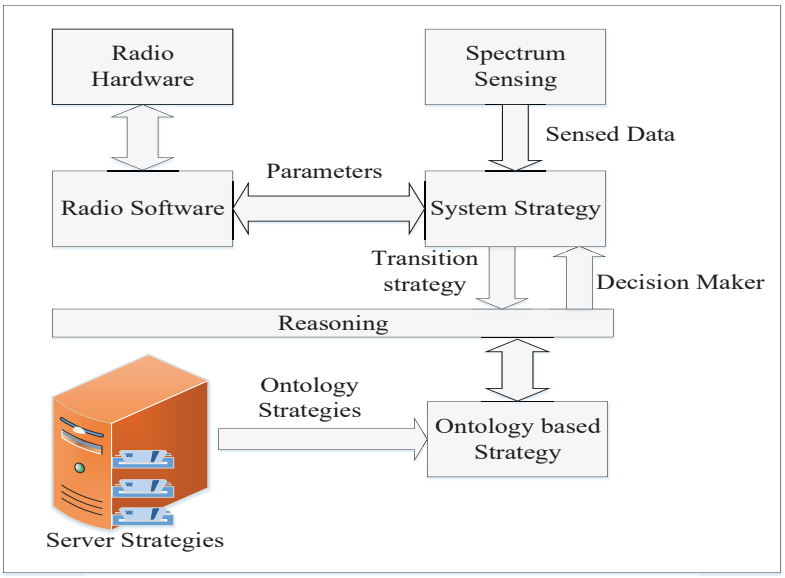


Figure 5.1: Design of experiment, adapted from (Bahrak et al., 2012)

5.2 Multi- Armed Bandit (MAB) Strategies

MAB strategies are employed as solutions for spectrum sensing and access problems, which is the dilemma of exploration and exploitation. Strategies are predicated on a gambler interacting with a number of slot machines and trying to be decide on which machine to play, gambling on a maximum return for minimal outlay, how many times are required to play this machine to meet the intended outcome and the playing arrangement for other machines. The basic form of this condition is expressed as:

Number of slot machines or arms = Number of armed bandits

MAB strategies requires many arms. Every arm, when played, has a single result. After each play, the result is checked against a sample of the result distribution ν_a, with probability ρ_a. Trying to discover the best arm to play, of the number of available arms K, at time t_i is expressed by equations 5.1 and 5.2.

$$a^* = argmax_a \rho_a \tag{5.1}$$

$$\rho^* = max_a \rho_a \tag{5.2}$$

Every machine produces rewards or varying amounts. The gambler's main goal is to boost the total rewards by making frequent decisions. In other words, the primary goal is to reduce regret R_T, defined as "The expected difference between the reward sum connected with an optimal strategy and the amount of the rewards collected under a particular strategy". The idea behind regret is that, if the gambler knew which is the best arm, that would be the only one played to increase chances of winning and maximizing rewards. But, lack of information affects the gambler by inevitable losses due to sub-optimal

plays. Equation 5.3 shows how regret is calculated (Chaczko et al., 2018).

$$R_T = T_{\rho^*} - \mathbb{E}\left[\sum_{t=1}^{T} X_t\right] \tag{5.3}$$

In the following sections, three MAB strategies to be applied in the simulations are identified and described.

5.3 Result and Discussion

Various analysis and scenarios have been performed to find "perfect" techniques of resolving the scarcity problem of available frequency bands in a wireless network's spectrum. These techniques depend on game theory and the adaption of MAB strategies to resolve the scarcity problem. Three scenarios for key MAB strategies have been applied and results were analyzed using multiple criteria to determine the best strategy. The best result maximizes rewards and reduces the number of regrets. The best result assists the identification of the best arm to play and thus resolves the bandwidth scarcity problem. The scenarios described in this section are discussed by contrasting results from the MAB strategies employed.

Scenario 1

The first scenario depends on the Bernoulli bandit system which assists the identification of the best arm to be played. Some random results from the first scenario's simulation were captured and displayed at Figure 5.1 - which illustrates the results of MAB strategies after applying the first scenario. As per figure 5.1, the strategies improved after changes were applied: noise decreased and the number of regrets were reduced. However, the major problem is the big difference between the three strategies. For better results lines, strategies over time should be closer to each other and asymptotic to the mean (shown as a solid line).

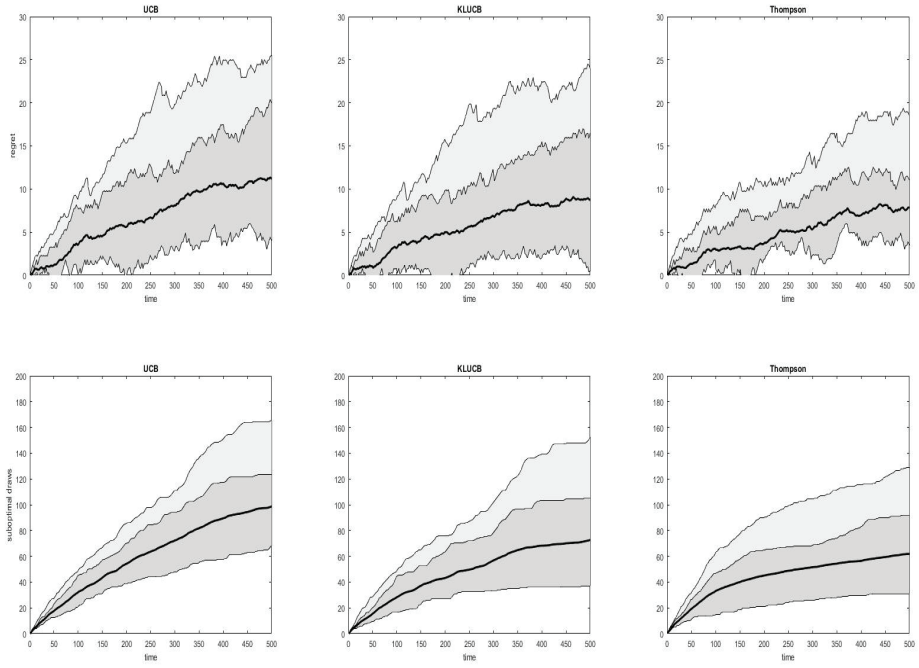


Figure 5.2: Results for scenario 1 MatLab Environment

Scenario 2

This scenario relies on a bounded exponential that assists an increase in the number of rewards and reduces the number of regrets. Figure 5.2 illustrate the results of the MAB strategies after applying this scenario. In the second scenario, noise was effectively reduced by comparison to the first scenario. When Thompson sampling was applied to the second scenario, it led to the least numbers of regrets, making it the better scenario to use (of scenarios 1 and 2).

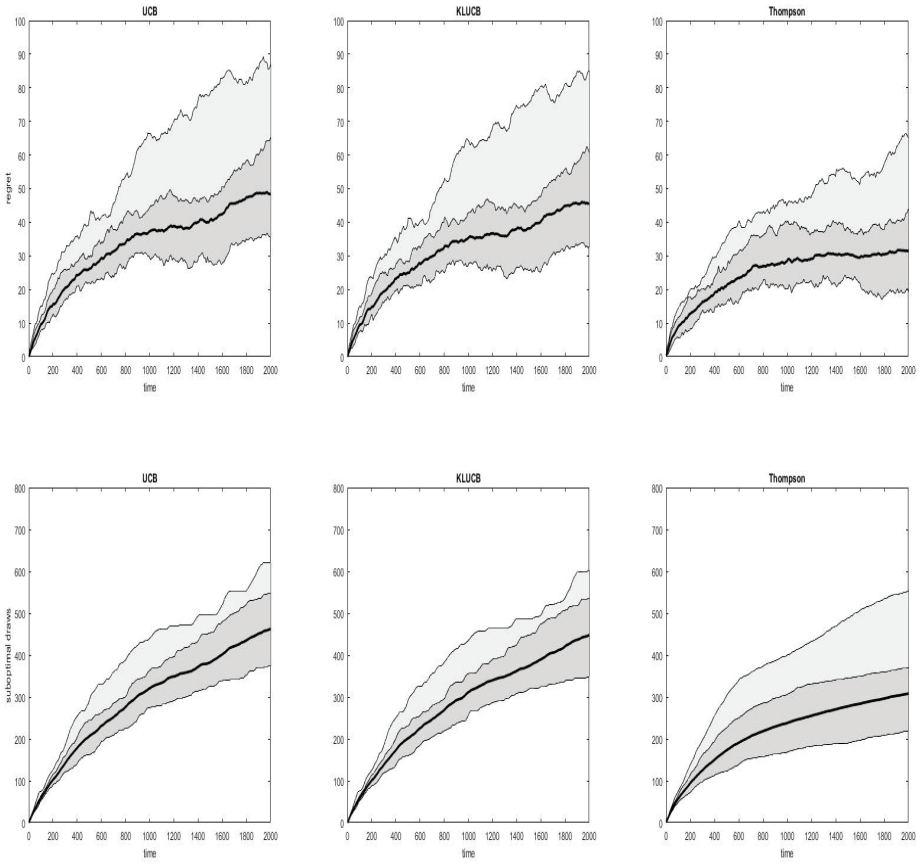


Figure 5.3: Results for scenario 2 MatLab environment

Scenario 3

The third scenario relies on bounded Poisson rewards, focusing on in-

creasing the number of rewards and reducing the number of regrets. Figure 5.3 illustrates the result of MAB strategies after applying the third scenario. Following the third scenario changes, noise has decreased much more effectively when compared to the first and second scenarios. Also, the lines become closer to the mean. When Thompson sampling was applied to the third scenario, it led to the least numbers of regrets than in the other two strategies, making it the better strategy to be employed.

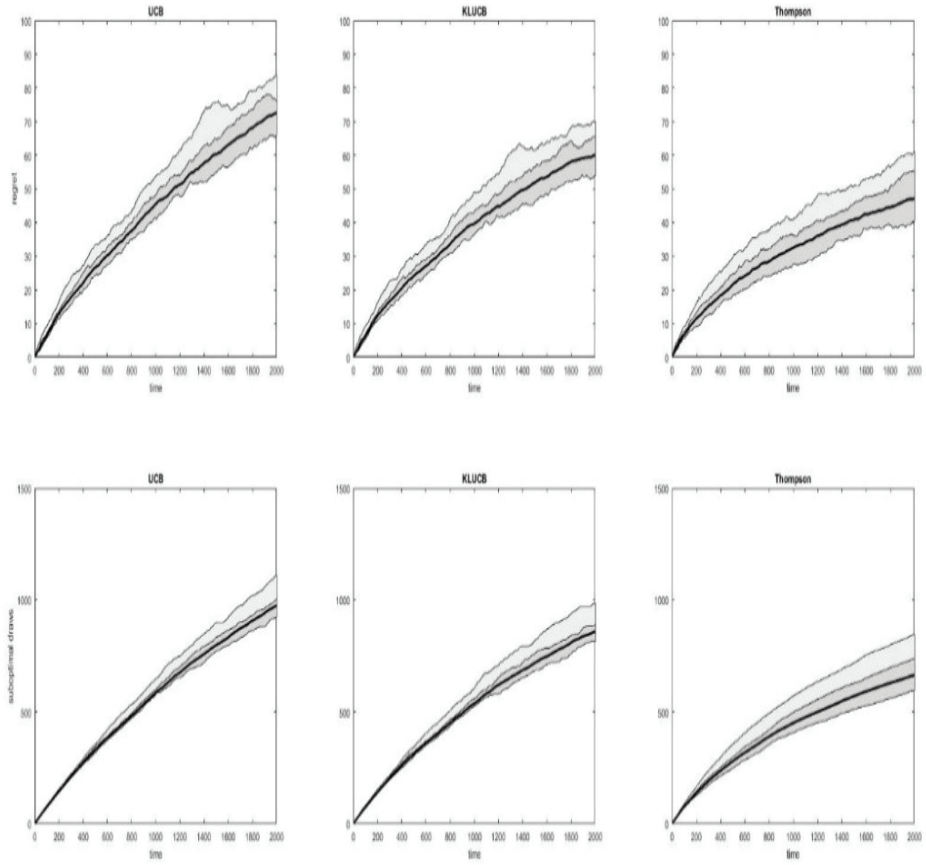


Figure 5.4: Results for scenario 3 MatLab environment

5.4 Upper Confident Bound

The bandit problem is a trade-off between exploration and exploitation. The main aim of Upper Confidence Bound (UCB) is to maximize rewards and reduce regrets. The conceptual UCB algorithm is shown at Algorithm 5.1 and its operation at Figure 5.2.

This research has studied frequencies simulations available for the CR. Using Python version 3.5.7 (Stewart, 2017), we have implemented versions of the Multi-Armed Bandits, under jamming strategies. The study began by the implementation of the Multi-Armed Bandits (UCB). Found through many simulations, the jammer going in any channel, we divide the jammer levels from 0 level to 5 levels. The red color is used to mark the jammer, blue to mark the algorithm, and brown to mark jammer and algorithm at the same time.

Algorithm 5.1 ***Upper Confident Bound***

1: **Begin**
2: **While** $t < 1$
3: $a^* = argmax_a \rho_a$
4: **end**
5: **While** $t \geq 1$
6: Compute point estimate = $\mu_i = R_i^t / T_i^t \forall i$
7: Compute index $g_i = \mu_i + \sqrt{\alpha \log \frac{t}{T_i^t}} \forall i$
8: Access channel $i^* = arg \max_i g_i$
9: Update $R_{i^*}^t$and $T_{i^*}^t$
10: **end**
11: **end**

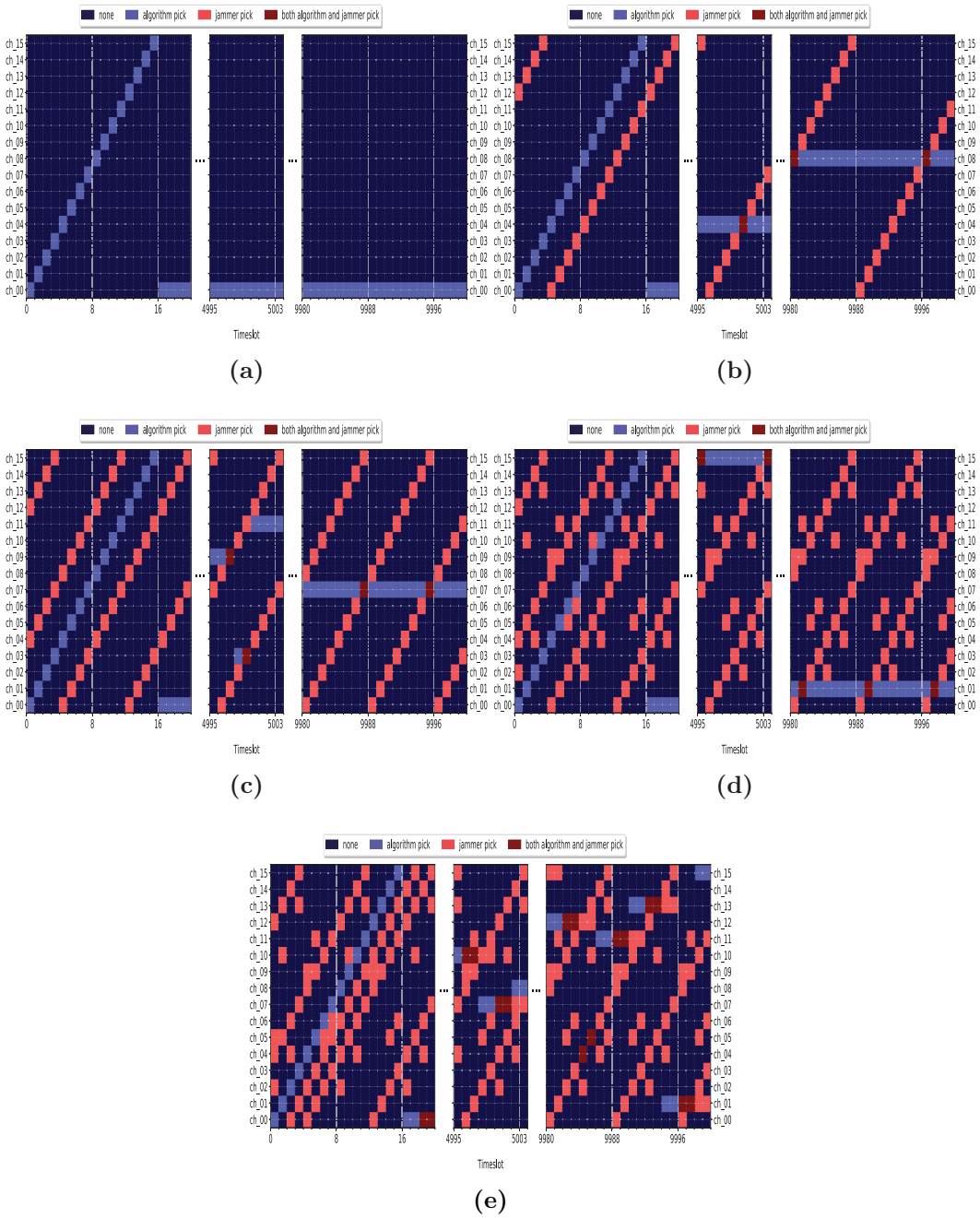


Figure 5.5: Environment with Jamming Level Zero Using Python Environment (Jupyter Notebook)

At level Zero: Picture (a) no jamming. Picture (b) jamming once in each channel, but in channel 4, and 8 jamming and UCB collide being applied to those channels at the same time. Picture (c) jamming

increases in all channels, but for channel 3, 7, 9 jamming and UCB collide being applied to those channels same time. Picture (d) increased jamming in all channels, but for channels 1, and 15 jamming and UCB collide being applied to those channels at the same time. (Picture (e) increased jamming in all channels, but for channels 1, 4, 5, 7, 10, 11, 12, 13 jamming and UCB occur in the same channels, but in channels 0, 1, 7, 10, 11, 12, 13 jammer and algorithm are in the same channel but in a different time slot. Figure 5.5 shows how the simulation behaves with conditions as shown above. Note: red color denotes the jammer, blue denotes the algorithm, and brown denotes the jammer and algorithm operating at the same time.

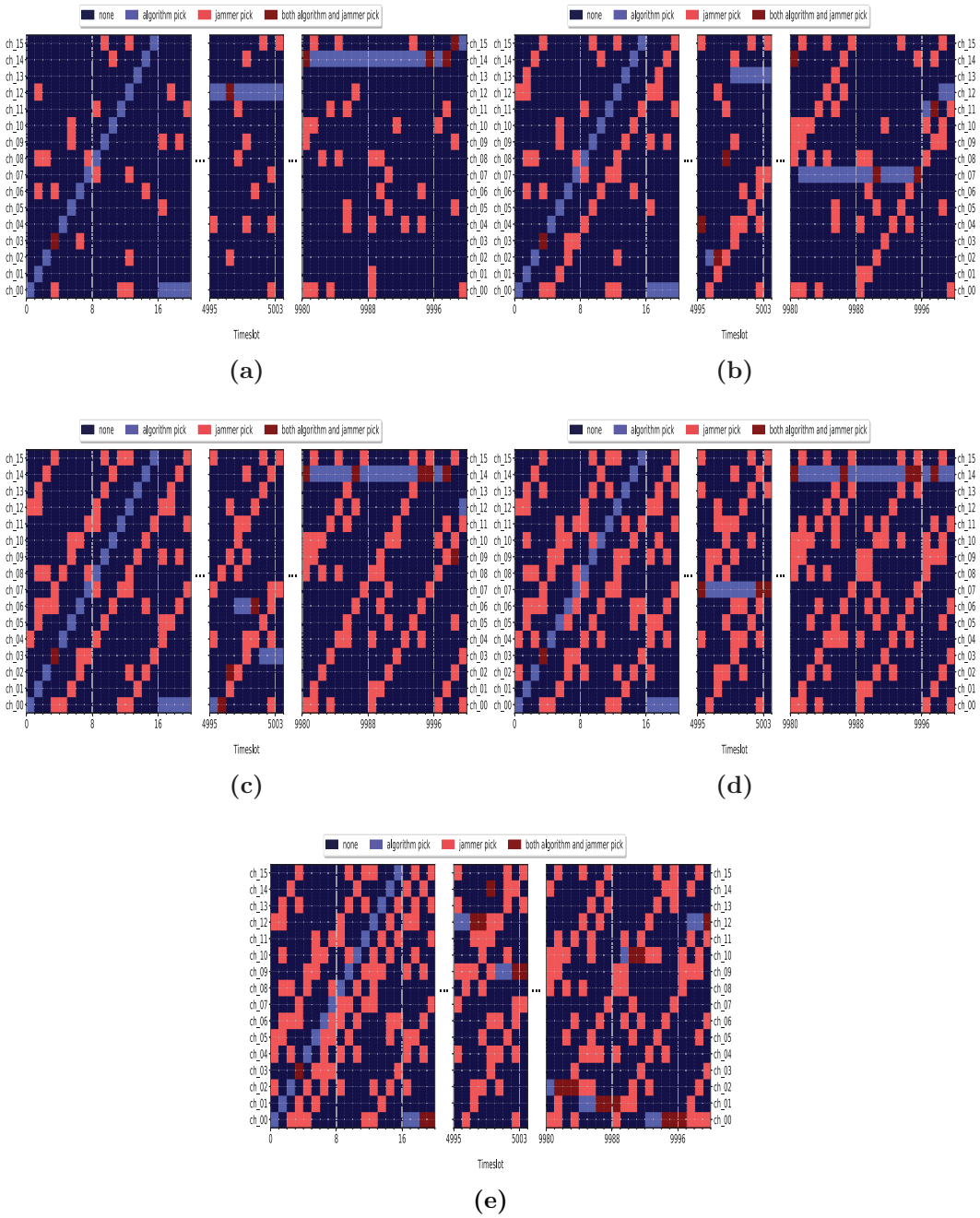


Figure 5.6: Environment with Jamming Level One Using Python Environment (Jupyter Notebook)

At level One, Picture (a) jamming randomly intrudes in channels, but in channel 3, 12, 14, 15, jamming and UCB collide being applied

to those channels at the same time, where as in channel 14 jamming and UCB are in the same channel but in a different time slot. Picture (b) jammer randomly in channels, but for channel 3, 7, 10 jammer and UCB collide being applied to those channels at the same time, channel 10, 11 jamming and UCB in the same channel but in a different time slot. Picture (c) jamming increases in all channels, but for channel 3, 4, 5, 7, 8, 11, 12, 14 jamming and UCB collide being applied to those channels at the same time, in channel 7, 11 jamming and UCB in the same channel but in a different time slot. Picture (d) increases jamming in all channels, but for channel 1, 2, 7, 10, 11, 12, 14 jamming and UCB collide being applied to those channels same time, in channel 7,14 jammer and algorithm same channel but in a different time slot. Picture (e) jamming increases in all channels, but for channel 0, 1, 2, 9, 10, 12, 14 jammer and UCB collide being applied to those channels at the same time, in channel 0,1, 2, 9, 10, 12 jammer and algorithm are in the same channel but in a different time slot. See figure 5.6.

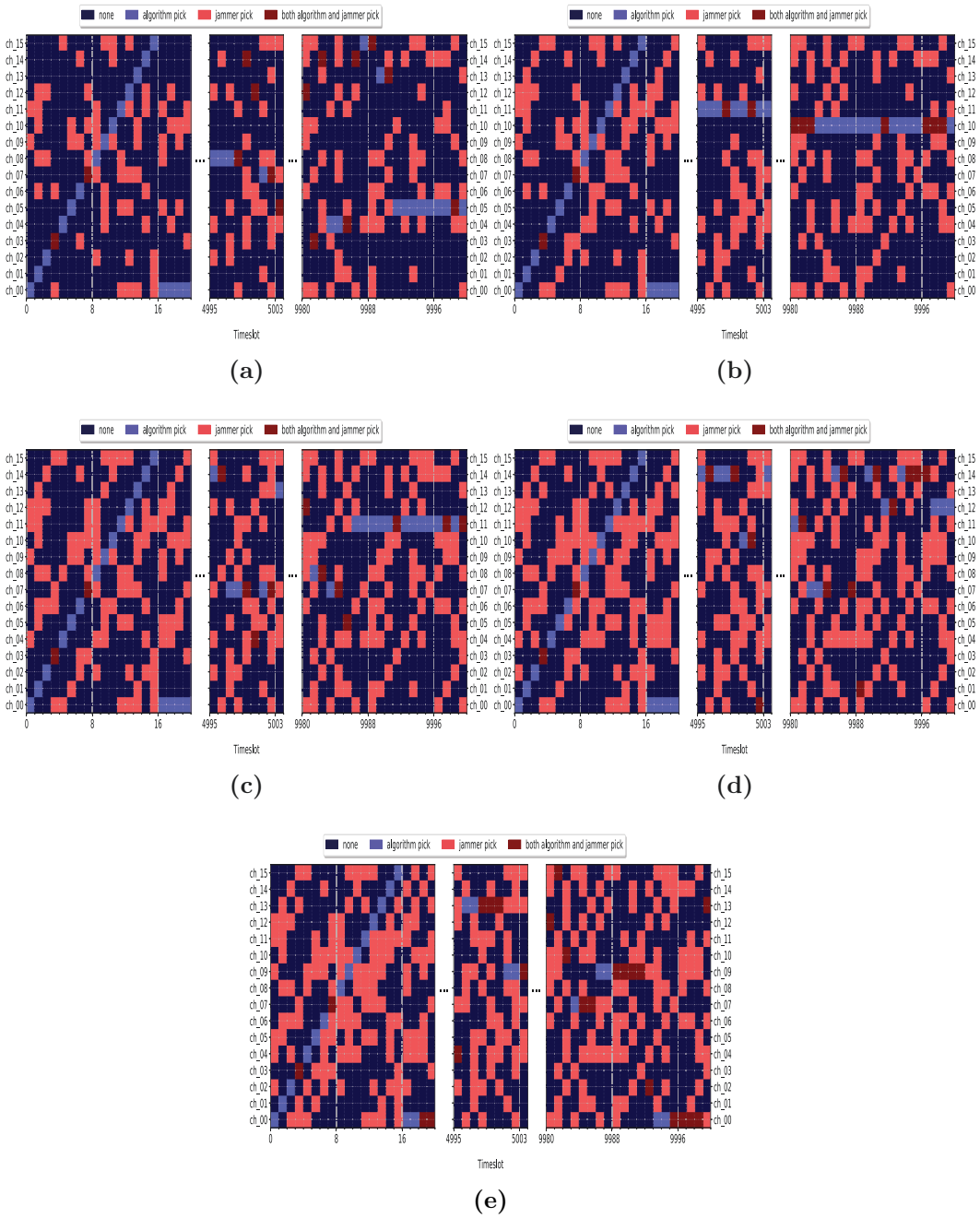


Figure 5.7: Environment with Jamming Level Two Using Python Environment (Jupyter Notebook)

At level Two, Picture (a) increases jamming in all channels, but for channels 3, 4, 5, 7, 8, 12, 13, 14, 15 jamming and UCB collide being

applied to those channels at the same time. Picture (b) increases jamming in all channels, but for channel 3, 7, 10, 11, jamming and UCB collide being applied to those channels at the same time, channel 10, 11, same channel but in a different time slot. Picture (c) increases jamming in all channels, but for channel 3, 4, 5, 7, 8, 11, 12, 14 jamming and UCB collide being applied to those channels at the same time, in channel 7, 11 jamming and USB at the same channel but in a different time slot. Picture (d) increases jamming in all channels, but for channel 1, 2, 7, 10, 11, 12, 14 jammer and UCB collide being applied to those channels at the same time, in channel 7,14 jamming and USB same channel but in different time slot. Picture (e) increases jamming in all channels, but for channel 0, 2, 3, 4, 7, 9, 10, 12, 13, 15 jamming and UCB collide being applied to those channels at the same time, in channel 0, 9, 13 jammer and algorithm same channel but in a different time slot. See figure 5.7.

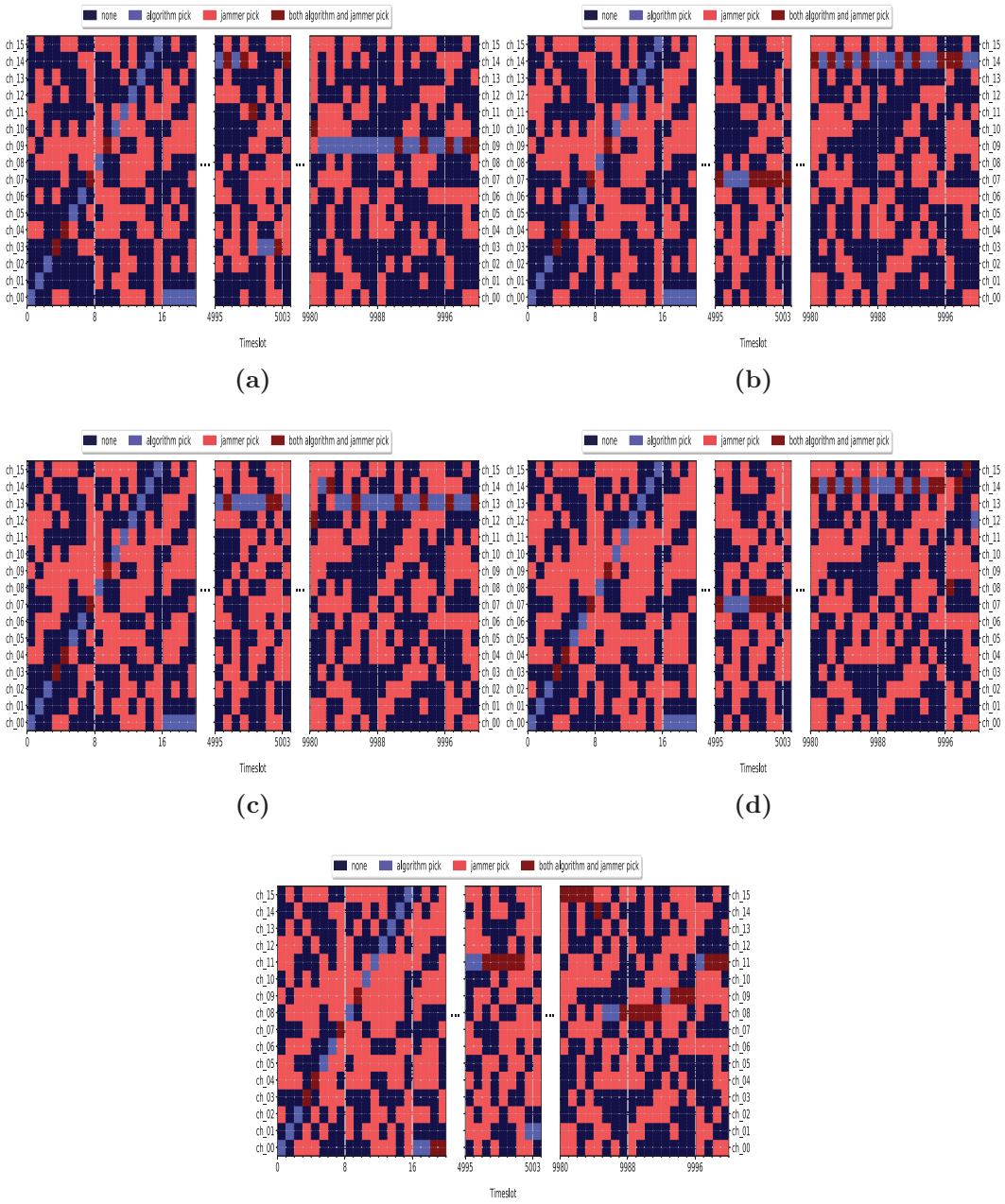


Figure 5.8: Environment with jamming Level Three Using Python Environment (Jupyter Notebook)

At level Three, Picture (a) jammer randomly in channels, but for channels, 3, 4, 7, 9, 10, 11, 14 jamming and UCB collide being applied

to those channels at the same time , in channel 3, 9, 14 jamming and USB in the same channel but in a different time slot. Picture (b) jammer randomly in channels, but for channel 3, 4, 7, 9, 14, jamming and UCB collide being applied to those channels same time, channel 7, 14, same channel but in different time slot. Picture (c) increases jamming in all channels, but for channel 3, 4, 7, 9, 12, 13, 14 jamming and UCB collide being applied to those channels at the same time, in channel 13 jammer and algorithm same channel but in different time slot. Picture (d) increases jamming in all channels, but for channel 3, 4, 7, 8, 9, 14, 15 jamming and UCB collide being applied to those channels at the same time, in channel 8, 15 jammer and algorithm same channel but in different time slot. Picture (e) increases jamming in all channels, but channel 0, 3, 4, 7, 8, 9, 10, 11, 14, 15 jamming and UCB collide being applied to those channels at the same time, in channel 0, 8, 9, 11, 15 jammer and algorithm same channel but in different time slot. See figure 5.8.

5.5 Kullback-Leibler Upper Confidence Bound (KLUCB)

KLUCB is dependent on the notion of gambling for a number of gamblers playing different machine arms for a particular time period to estimate the “best arm”. KLUCB uses the results of “best arm” selec-

tion to determine the optimal Maximum Point Estimate (MPE) (Lai and Robbins, 1985). The KLUCB algorithm is shown at algorithm 5.2.

Algorithm 5.2 ***KLUCB Algorithm*** **(Gwon et al., 2013)**

1: **Begin**
1: **While** $t < 1$
2: Access each channel at least once
3: Record $R_i^t = \sum_{j=1}^{t} r_i^j$ and T_i^t for every channel$i$
4: **end**
5: **While** $t \geq 1$
6: Calculate $\mu_i = R_i^t / T_i^t \ \forall_i$
7: Find MPE candidate $C_{MPE} = i^*$ s.t $\mu_{i^*} = \max \mu_i$
8: Find RR candidate $C_{RR} = (t \bmod N) + 1$
9: If $D_{KL}(P_{RR} \| P_{MPE}) > \log(t - 1) / T_{C_{RR}}^t$
10: Access $C_{MPE}$and monitor $r_{C_{MPE}}^t$
11: Update $R_{C_{MPE}}^t$ and $T_{C_{MPE}}^t$ observe T_{MPE}^t
12: **else**
13: Access C_{RR} and monitor $r_{C_{RR}}^t$
14: Update $R_{C_{RR}}^t$ and $T_{C_{RR}}^t$
15: **end**
16: **end**

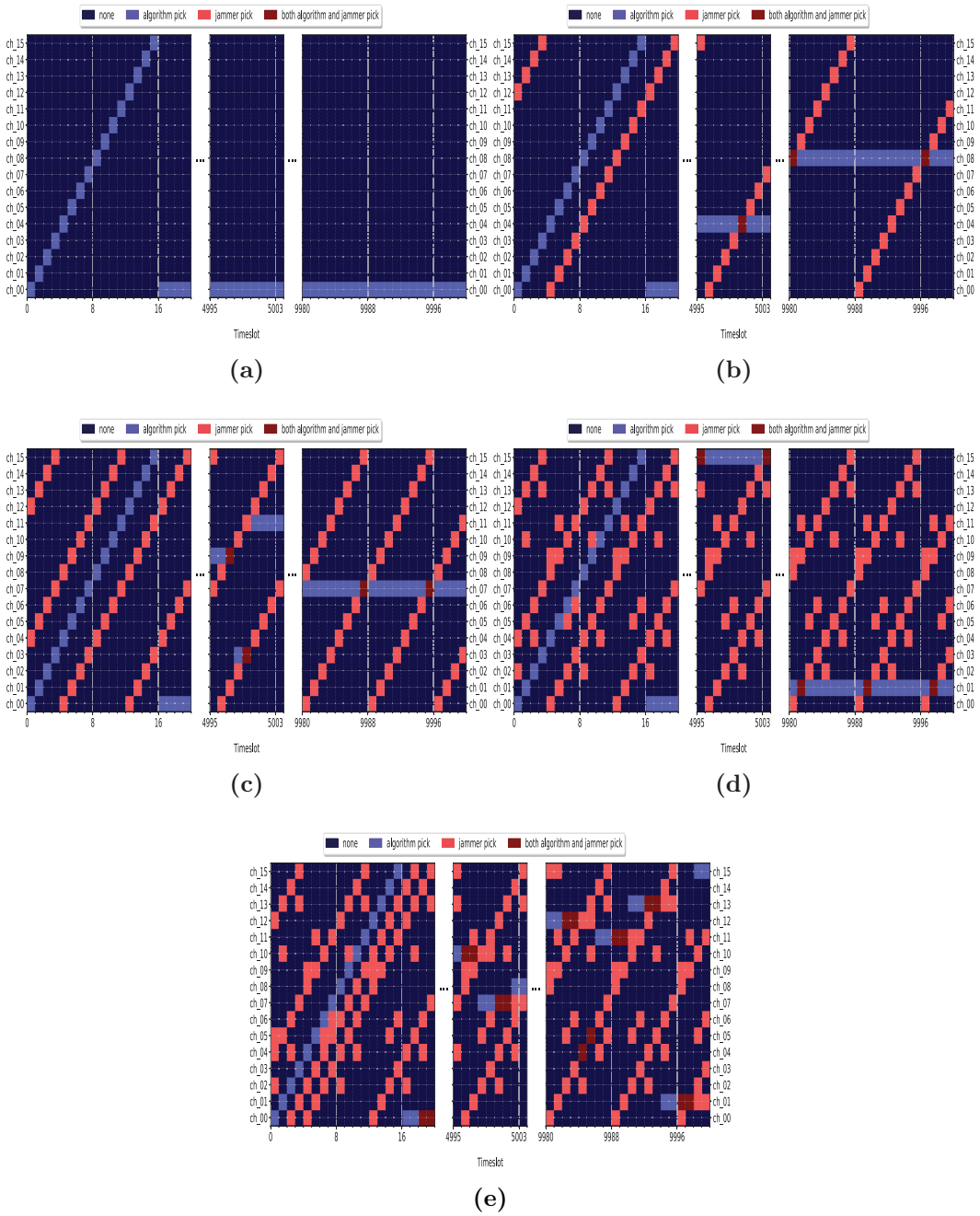


Figure 5.9: Environment with Jamming Level Zero Using Python Environment (Jupyter Notebook)

At level Zero, Picture (a) no jam at all. Picture (b) the jamming is in each channel, but for channel 4, and 8 jamming and KLUCB collide

being applied to those channels at the same time. Picture (c) increases jamming in all channels, but for channel 3, 7, 9 jamming and KLUCB collide being applied to those channels at the same time. Picture (d) increases jamming in all channels, but for channel 1, and 15 jamming and KLUCB collide being applied to those channels at the same time. Picture (e) increases jamming in all channels, but for channel 0, 1, 1, 4, 5, 7, 10, 11, 12, 13 jamming and KLUCB collide being applied to those channels at the same time, in channel 0, 1, 4, 5, 7, 10, 11, 12, 13 jamming and KLUCB in the same channel but in different time slot. See Figure 5.9 for the result of the algorithm against jamming.

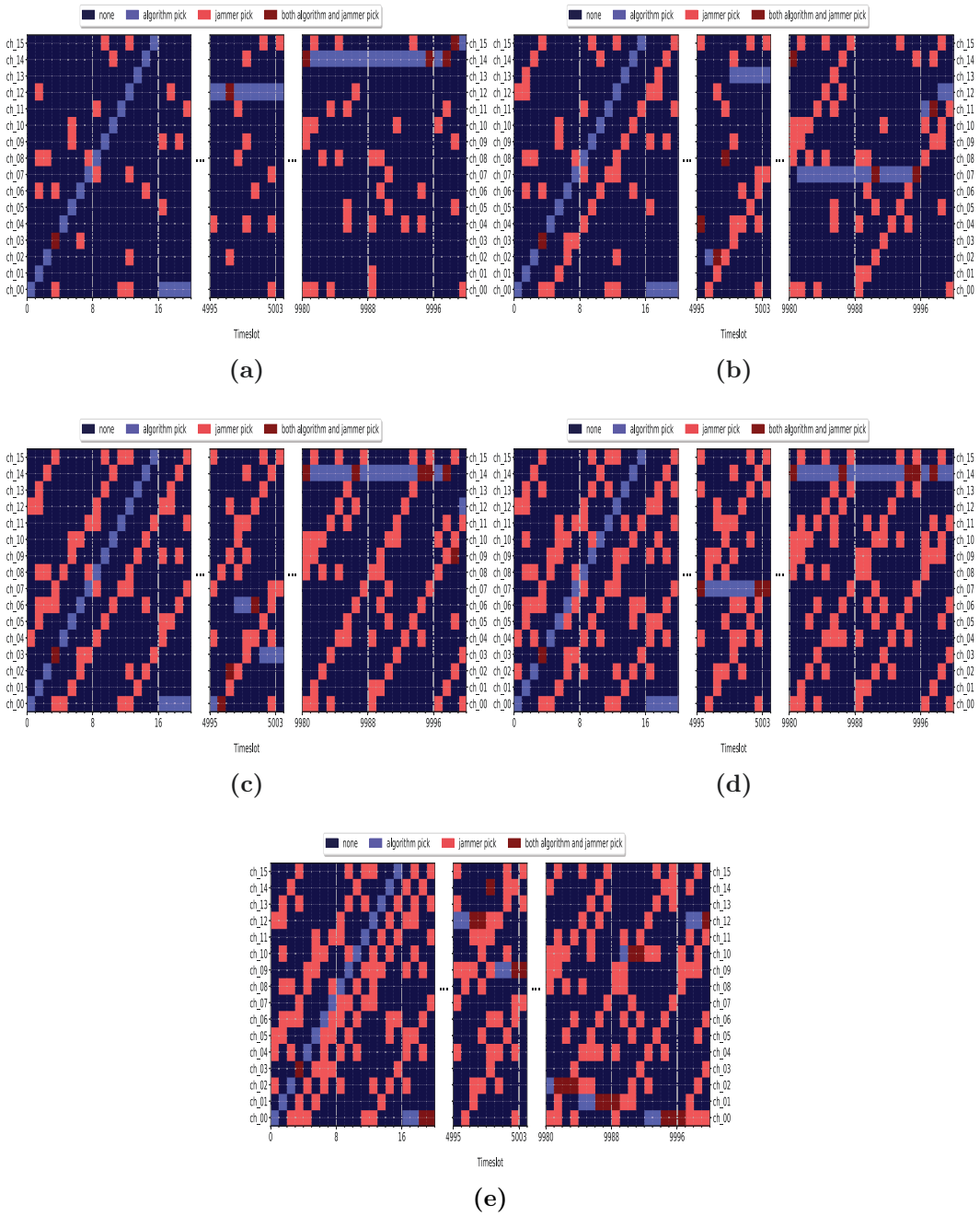


Figure 5.10: Environment with Jamming Level One Using Python Environment (Jupyter Notebook)

At level One, Picture (a) jamming randomly intrudes in channels, but for channel 3, 12, 14, 15 jamming and KLUCB collide being ap-

plied to those channels at the same time, whereas in channel 14 jamming and KLUCB same channel but in different time slot. Picture (b) jamming randomly intrudes in channels, but for channel 2, 4, 7, 8, 11, 14 jamming and KLUCB collide being applied to those channels at the same time, channel 7 jamming and KLUCB collide but in a different time slot. Picture (c) increases jamming in all channels, but for channel 0, 2, 3, 6, 14 jamming and KLUCB collide being applied to those channels at the same time, in channel 14 jammer and KLUCB in the same channel but in different time slot. Picture (d)increases jamming in all channels, but for channel 7, 14 jamming and KLUCB collide being applied to those channels at the same time and also in different time slot. Picture (e) increases jamming in all channels, but for channel 0, 1, 2, 3, 9, 10, 12, 14 jamming and KLUCB collide being applied to those channels at the same time and also in different time slot, channel 0, 1, 2, 9, 12 jammer and KLUCB collide but in different time slot. See figure 5.10

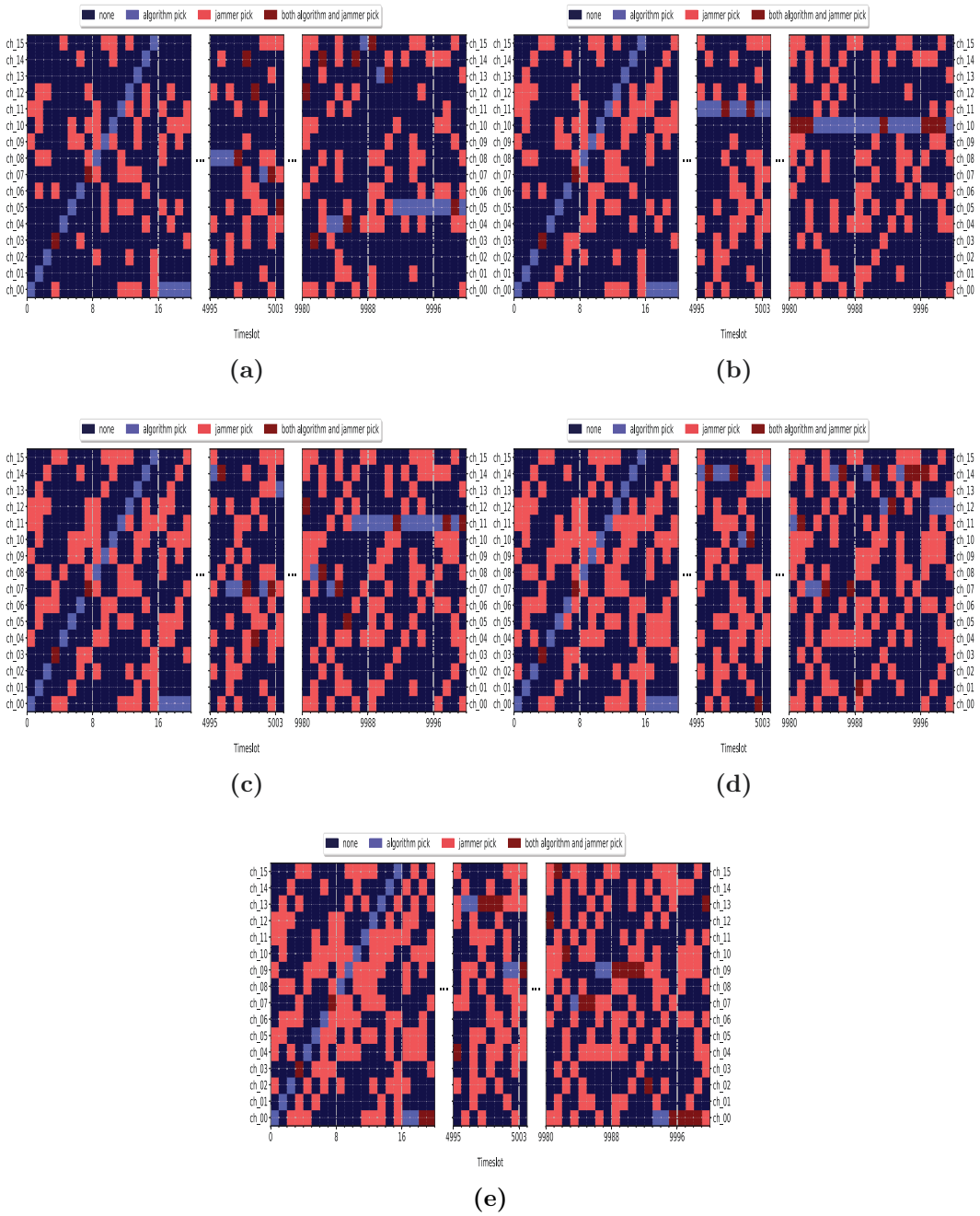


Figure 5.11: Environment with Jamming Level Two Using Python Environment (Jupyter Notebook)

At level Two, Picture (a) jammer randomly intrudes in channels, but for channel 3, 4, 5, 7, 8, 12, 13, 14, 15 jamming and KLUCB

collide being applied to those channels at the same time, whereas in channels 5, 12, 14 jamming and KLUCB in the same channel but in a different time slot. Picture (b) increases jamming in all channels, but for channel 3, 7, 10, 11 jamming and KLUCB collide being applied to those channels at the same time, channel 10, 11 jamming and KLUCB collide but in a different time slot. Picture (c) increases jamming in all channels, but for channel 3, 4, 5, 7, 8, 11, 12, 14 jamming and KLUCB collide being applied to those channels at the same time. In channel 7, 14 jamming and KLUCB collide but in a different time slot. Picture (d) increases jamming in channels, but for channel 0, 1, 7, 10, 11, 12, 14 jamming and KLUCB collide being applied to those channels at the same time, in channel 7, 10, 14 jamming and KLUCB collide but in a different time slot. Picture (e) increases jamming in all channels, but for channel, 0, 2, 4, 7, 9, 10, 12, 13, 15 jamming and KLUCB collide being applied to those channels at the same time, in channel 0, 2, 9, 13 jamming and KLUCB collide but in a different time slot. See figure 5.11.

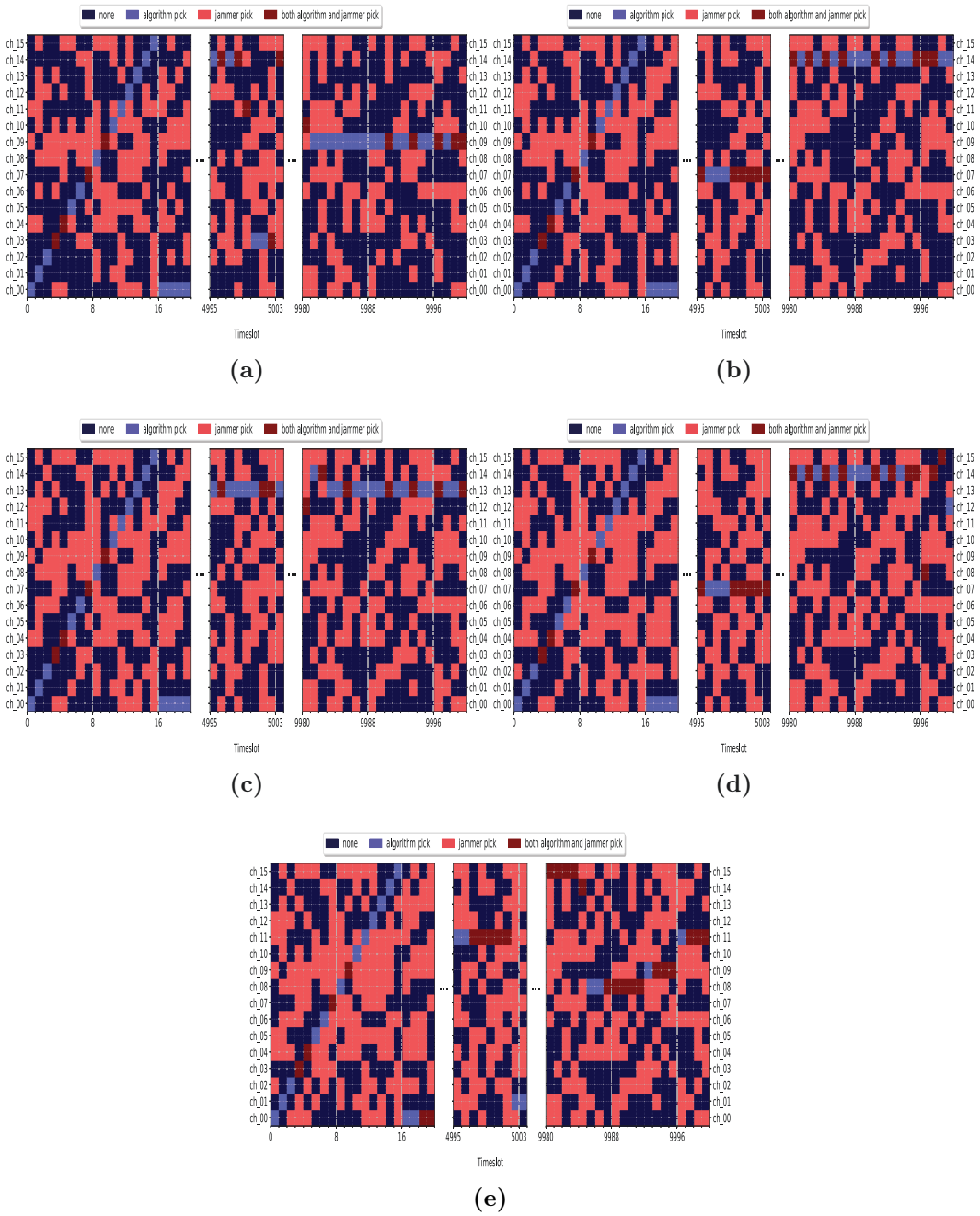


Figure 5.12: Environment with Jamming Level Three Using Python Environment (Jupyter Notebook)

At level Three, Picture (a) jammer randomly intrudes in channels, but for channel 3, 4, 7, 9, 11, 14 jamming and KLUCB collide being

applied to those channels at the same time, whereas in channels 9, 14 jamming and KLUCB are in the same channel but in a different time slot. Picture (b) jamming randomly intrudes in channels, but for channel 3, 4, 7, 9, 14 jamming and KLUCB collide being applied to those channels same time, whereas in channel 7, 14 jamming and KLUCB collide but in a different time slot. Picture (c) increases jamming in all channels, but for channel 3, 4, 7, 9, 12, 13, 14 jamming and KLUCB collide being applied to those channels at this same time, whereas in channel 13 jamming and KLUCB collide but in a different time slot. Picture (d) increase jamming in all channels, but for channel 3, 4, 7, 9, 14, 15 jamming and KLUCB collide being applied to those channels at the same time, whereas in channel 7, 15 jammer and KLUCB collide but in a different time slot. Picture (e) increases jamming in all channel, but for channel 0, 3, 4, 8, 9, 10, 11, 14, 15 jamming and KLUCB collide being applied to those channels at the same time, whereas channel 0, 8, 11, 15 jamming and KLUCB collide but in a different time slot. See figure 5.12.

5.6 Thompson Sampling (TS)

The Thompson strategy using sampling and probability matching, originating in the 1930s, is specifically used in experiments to solve

the two armed bandits problem (Wyatt, 1998; Strens, 2000). In recent years the Thompson strategy has attracted a large amount of literature and has been successfully applied to many applications.

The strategy is applied as follows: suppose that the reward for playing each arm is created from the variable distribution ν_i (Korda et al., 2013). Then the structure of the algorithm includes:

- For each arm, begin with a prior confidence, on the distribution variables .
- On making observations from an arm, update to posterior belief.
- At time t the arm that is played is selected from a calculated probable desired arm.

The Thompson sampling algorithm is shown at Algorithm 5.3

Algorithm 5.3 Thompson Sampling Algorithm (Gwon et al., 2013)

1: **Begin**
2: Require: $d = \{X, a, r\}$ for X, action a, reward r, estimator $p(\theta \mid d)$ α
3: $p(r \mid x, a, \theta)p(\theta)$ parameterized by θ
4: **While** $t \geq 1$
5: Acquire x^t
6: Draw $\theta^t \sim p(\theta)$
7: Select a^t to access $i^* = arg\max_i \mathbb{E}\left[r_i^t \mid x^t, \theta^t\right]$
8: Observe actual r^t
9: Update $d = d \cup \{x^t, a^t, r^t\}$
10: Update $p(\theta) = p(\theta \mid d)$
11: **end**
12: **end**

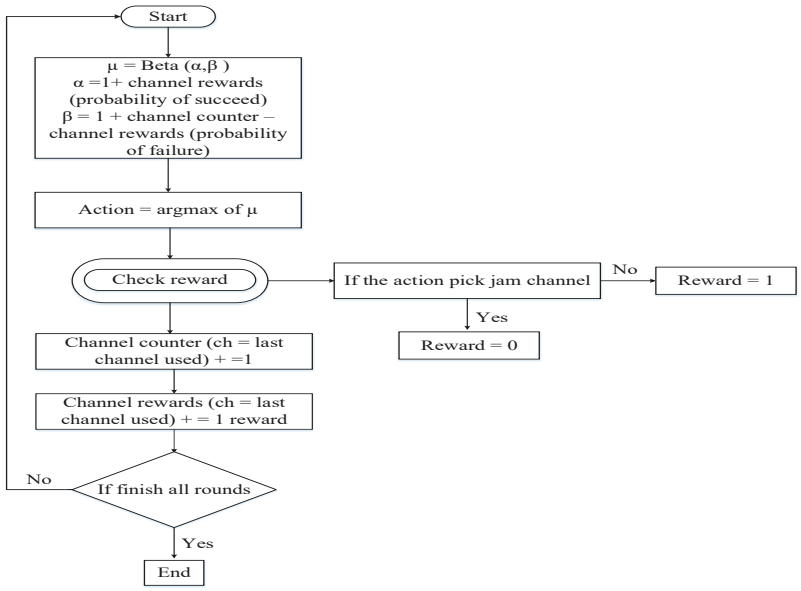


Figure 5.13: Flowchart of the Thompson Sampling Process

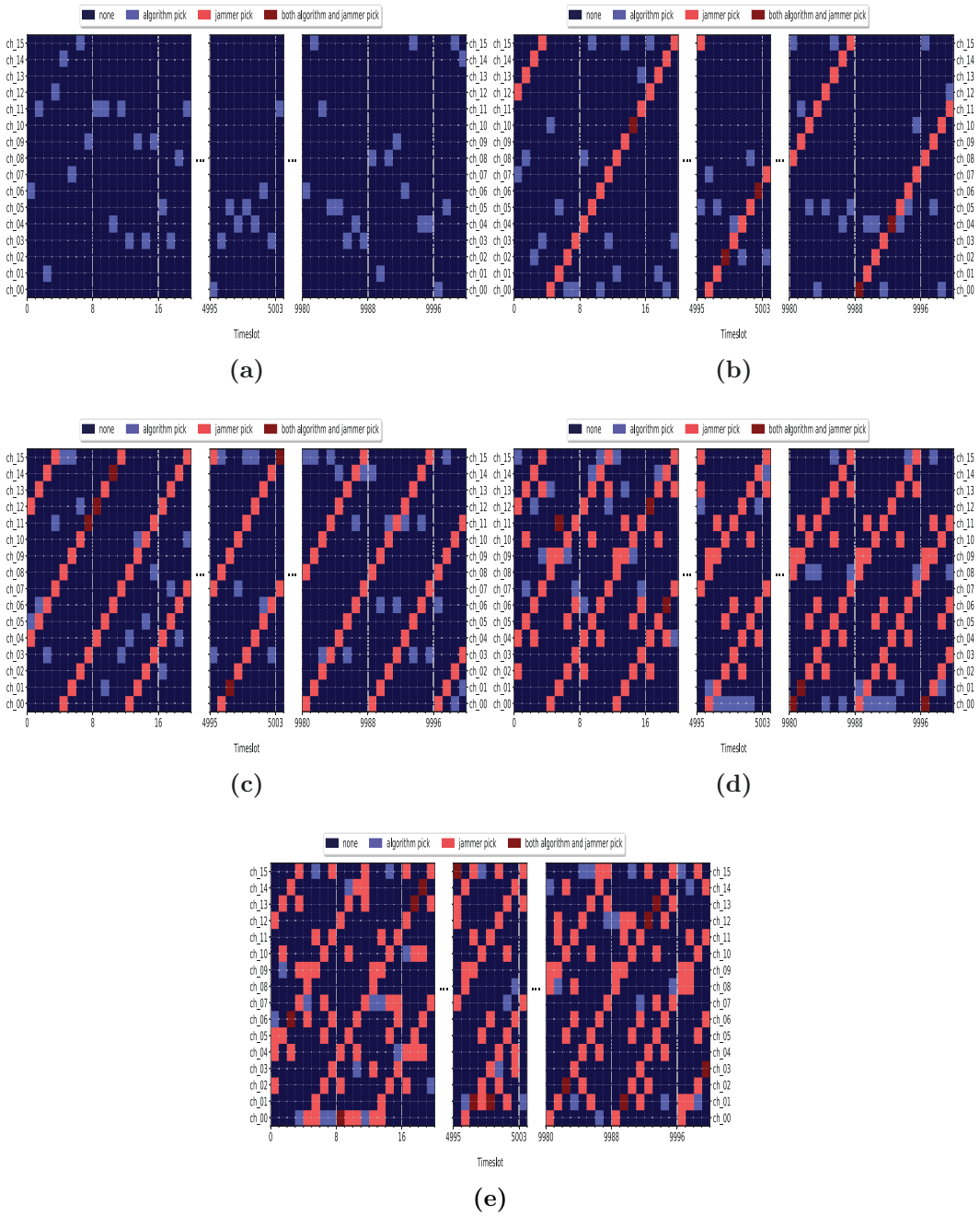


Figure 5.14: Environment with Jamming Level Zero Using Python Environment (Jupyter Program)

At level Zero, Picture (a) no jam at all. Picture (b) jamming intrudes in channels, but for channel 1, 11, 14, 15 jamming and TS

collide being applied to those channels at the same time. Picture (c) increases jamming in all channels, but for channel 1, 11, 14, 15 jamming and TS collide being applied to those channels at the same time. Picture (d) jamming intrudes in channels, but for channel 0, 1, 6, 11, 12 jamming and TS collide being applied to those channels at the same time. Picture (e) increases jamming in all channels, but for channel 0, 1, 2, 3, 6, 12, 13, 14, 15 jamming and TS collide being applied to those channels at the same time, whereas in channels 1, 13 jamming and TS are in the same channel but in a different time slot. Figure 5.14 shows how the algorithm against jamming is working.

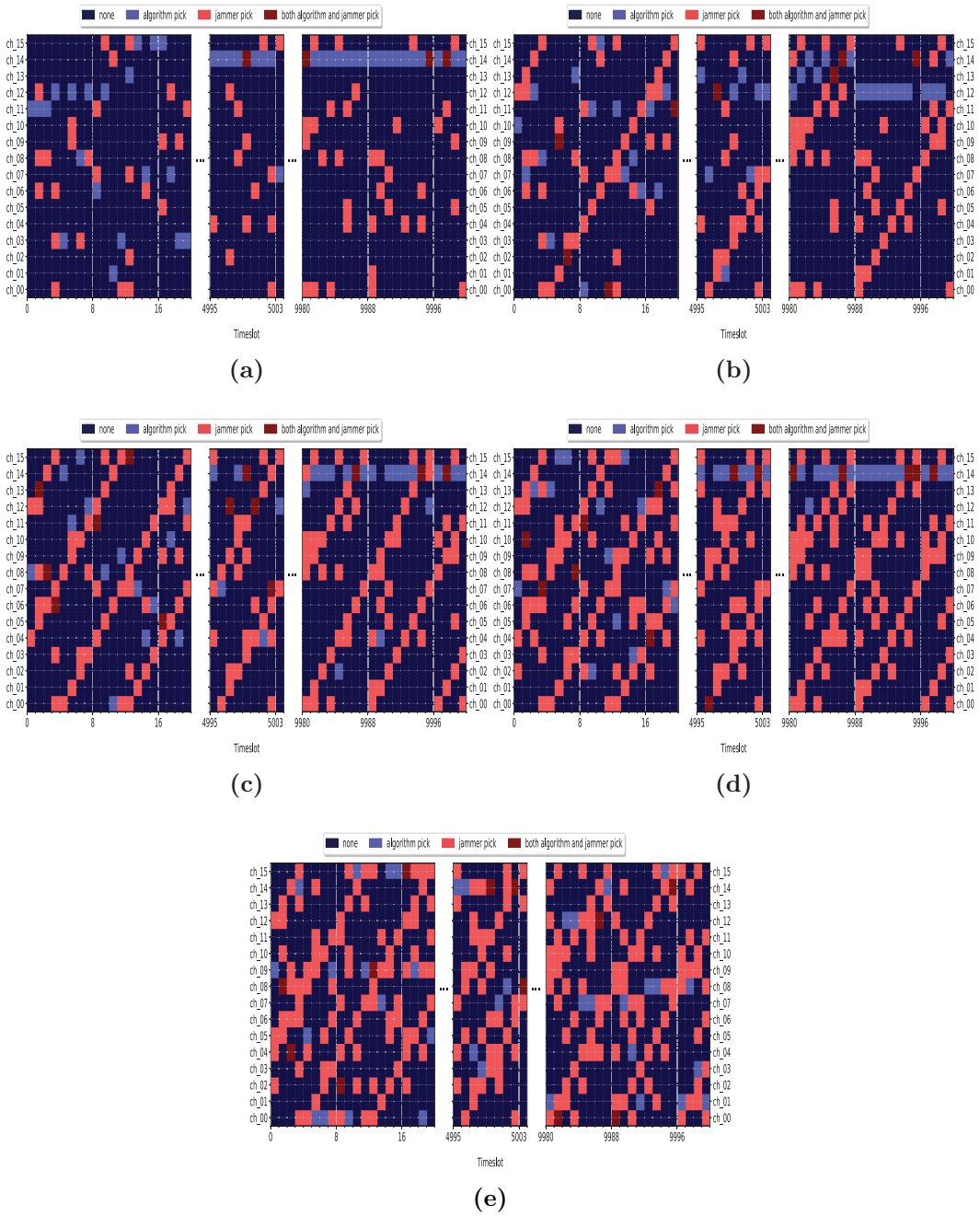


Figure 5.15: Environment with Jamming Level One Using Python Environment (Jupyter Notebook)

At level One, Picture (a) jammer randomly intrudes in channels, but for channel 14 jamming and TS collide being applied to those channels

at the same time, whereas in channel 14 jamming and algorithm same channel but in a different time slot. Picture (b) increase jamming in all channels, but for channel 0, 2, 9, 7, 11, 12, 13, 14 jamming and TS collide being applied to those channels at the same time. Picture (c) increases jamming in all channels, but for channel 5, 6, 7, 8, 11, 12, 13, 14, 15 jamming and TS collide being applied to those channels at the same time, whereas in channel 12, 14 jamming and TS collide but in a different time slot. Picture (d) increases jamming in all channel, but for channel 0, 4, 7, 8, 10, 11, 12, 13, 14 jamming and TS collide being applied to those channels at the same time, whereas in channel 14 jamming and TS collide but in a different time slot. Picture (e) increases jamming in all channels, but for channel 0, 2, 4, 8, 9, 10, 12, 14, 15 jamming and TS collide being applied to those channels at the same time, whereas in channel 0, 1, 8, 14 jamming and TS collide but in a different time slot. Figure 5.15 shows the environment with jamming level one.

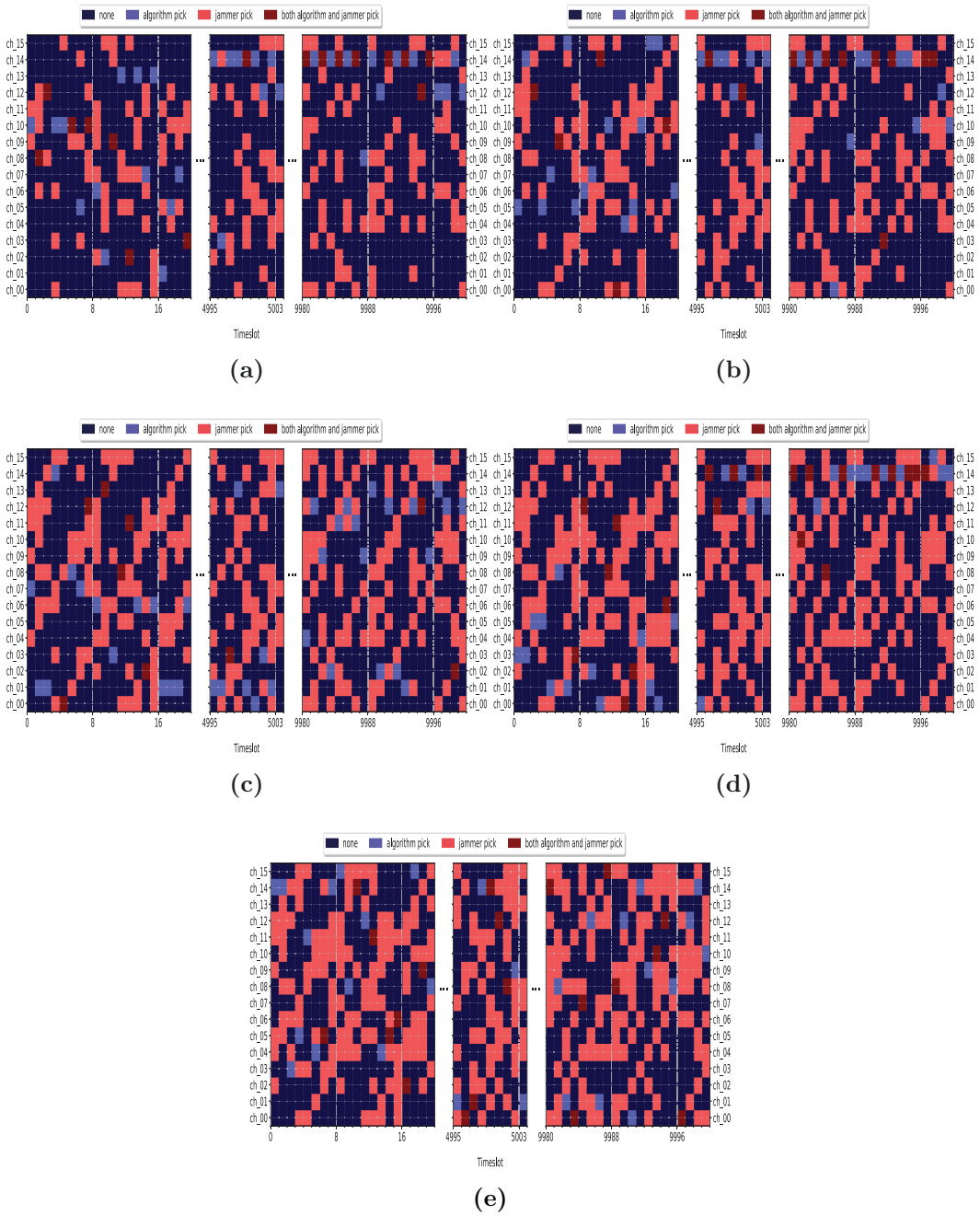


Figure 5.16: Environment with Jamming Level Two Using Python Environment (Jupyter Notebook)

At level Two, Picture (a) jammer randomly intrudes in channels, but in channels 2, 3, 8, 9, 10, 12, 14 jamming and TS collide being

applied to those channels at the same time, whereas in channel 10, 12, 14 jamming and TS collide but in a different time slot. Picture (b) increases jamming in all channels, but for channel 0, 3, 9, 10, 12, 14 jamming and TS collide being applied to those channels at the same time, whereas in channel 12, 14 jamming and TS collide but in a different time slot. Picture (c) increases jamming in all channels, but for channel 0, 2, 3, 8, 11, 12, 13 jammer and TS collide being applied to those channels at the same time, whereas in channel 2, 12, jammer and TS collide but in a different time slot. Picture (d) increases jamming in all channels, but for channel 0, 2, 3, 6, 8, 10, 12, 14 jamming and TS collide being applied to those channels at the same time ,whereas in channel 8, 14 jammer and TS collide but in a different time slot. Picture (e) increases jamming in all channels, but for channel 0, 1, 2, 5, 7, 8, 9, 10, 11, 12, 14, 15 jamming and TS collide being applied to those channels at the same time, whereas in channel 0, 5, 8, 12, 14 jammer and TS collide but in a different time slot. Figure 5.16 shows the environment with jamming level two.

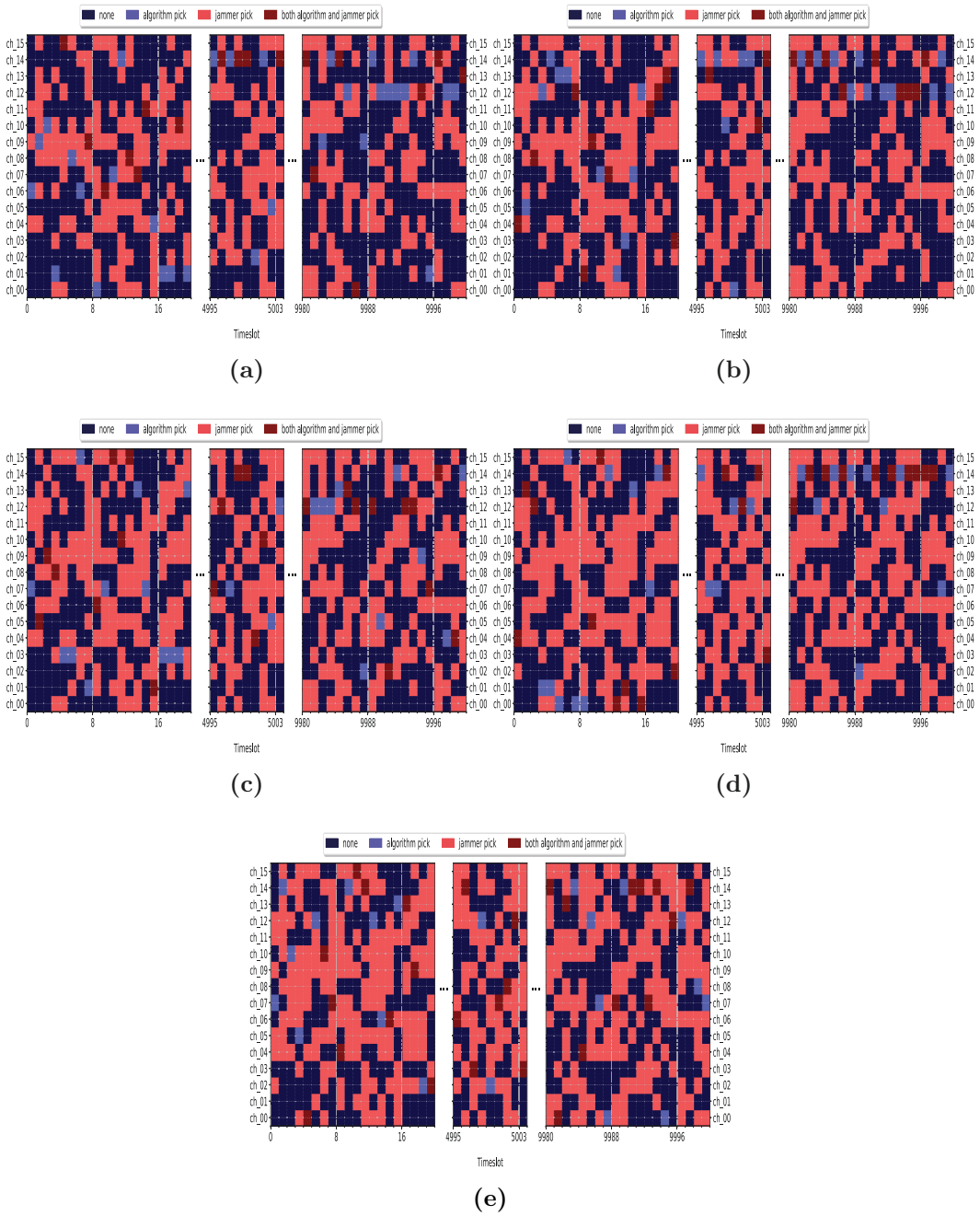


Figure 5.17: Environment with Jamming Level Three Using Python Environment (Jupyter Notebook)

At level Three, Picture (a) jamming randomly intrudes in channels, but for channel 0, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 jamming and TS

collide being applied to those channels at the same time, whereas in channel 7, 14 jamming and TS collide but in a different time slot. Picture (b) jamming increases in all channels, but for channel 1, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14 jamming and TS collide being applied to those channels at the same time, whereas in channel 12, 13, 14 jammer and algorithm collide but in a different time slot. Picture (c) increases jamming in all channels, but for channel 1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15 jamming and TS collide being applied to those channels at the same time, whereas in channel 10, 12, 14, 15, jamming and TS collide but in a different time slot. Picture (d) increases jamming in all channels, but for channel 0, 1, 2, 3, 4, 5, 12, 13, 14, 15 jamming and TS collide being applied to those channels same time, whereas in channel 0, 12, 14 jamming and TS collide but in different time slot. Picture (e) increases jamming in all channels, but for channel 0, 2, 3, 4, 7, 8, 9, 10, 12, 14, 15 jamming and TS collide being applied to those channels at the same time, whereas in channel 0, 3, 4, 6, 7, 8, 9, 10, 12, 14, 15 jammer and TS collide but in a different time slot. Figure 5.17 shows the environment with jamming level two.

5.7 ProML: A Method for Cognitive Radio Jamming Attack Simulation and Protection Using Machine Learning Approach

Machine learning algorithms have largely influenced information engineering and IT applications due to their ability to learn from the past and predict future events. In the case of Cognitive Radio security, machine-learning algorithms can provide a robust solution, based on learning (via signal parameters) about breaches while they are occuring. Major areas to be addressed by machine learning for CR include:

- Detecting unusual events,
- Avoiding channel blockages,
- Predicting future trends.

This section shows the implementation of a strategy based on 'Random-Forests' algorithms, to provide a dynamic selection mechanism to overcome random attacks on CR radio channels. This method is called (ProML), which is the acronym of 'Protection using Machine Learning'. This section includes five subsections: background of Machine Learning methods as applied in Cognitive Radio security; a proposed new method versus other conventional approaches; a simulation using synthetic data-sets, which imitates the operation of real Cognitive Radio signals in case of jamming, idle, busy and conges-

tion scenarios; discussion on merits and shortfalls of the proposed new method and conclusion.

5.7.1 Background

This subsection provides a very quick review of the Machine Leaning methods used for anti-jamming application in Cognitive Radio security. Three major aspects of the security including unusual event detection, blockage avoidance and future prediction are shown and discussed:

- **Unusual Event Detection using Machine Learning Methods**:

(Slimeni et al., 2015) proposed a method in which the use of Q-learning determines anti-jamming strategies to avoid jammed channels, proactively. The issue with Q-learning is that it requires a large training time to learn from the actions of the jamming. To mitigate this, Q-learning uses the wide band spectrum sensing abilities of Cognitive Radio to speed up the learning process. The method also takes advantage of already learned information to reduce the number of collisions between legitimate transmissions and jamming, by training the method. Simulations assess the efficiency of the method in the face of four jamming strategies and the simulation outcomes are compared

with the original Q-learning algorithm implemented to the scenarios described below.

(Machuzak and Jayaweera, 2016) proposed a reinforcement learning approach for overcoming accidental and purposeful jamming in Cognitive Radios. In this method, the cumulative rewarding is in reinforcement learning as used for determining the optimal communication mode to avoid jamming. The authors developed a scheme called Wide Band Autonomous Cognitive Radio (WACR) to mitigate jamming. (Ling et al., 2015), proposed an approach which focuses on use of reinforcement learning for detecting malicious nodes. This has the capability to detect new attacks and record data from previous attacks. This approach has proved to be a useful technique to enhance security and predict future attacks.

- **Mitigating Wireless Jamming Attacks**

A cognitive transmitter utilities a per-trained classifier to predict existing channel statuses dependent on current sensing outcomes and decisions on whether to transmit or not. Meanwhile the jammer gathers states of channels and acknowledgments to establish a deep learning classifier that predicts the completion of the next transmission. This targeted jamming approach has been shown to minimize a transmitter's performance much more than random or sensing based jamming.

The machine learning classifier is used by a jammer to control energy according to medium power constraints. A cumulative mixing network was devised to minimize the time required to collect a training data-set by increasing its size using synthetic data. As a defence schema, the transmitters purposely make a small number of faulty transmissions in the spectrum preventing the jammer from establishing a trusted stable classifier. The transmitter regularly chooses when to select wrong actions and adapts its level of defence to cheat the jammer into giving predicting errors and consequently enhancing its own throughput.

5.7.2 ProMl Approach

The ProMl approach focuses on combining advantages of machine learning methods that have been discussed previously. The main objective is to develop a method which is both high performing and adaptive to jamming attacks. Various stages of this approach include: training and simulation of events, validating state detections and testing. Simulation of events is developed using signal parameter data, which represents every state. This data is developed using prior knowledge determined from historical sampling. (Machuzak and Jayaweera, 2016) showed signal parameters plus their behaviour in five main features, as shown below. Empirical evidence suggests that the following

features of Cognitive Radio may be affected due to busy, idle, congested and jammed states:

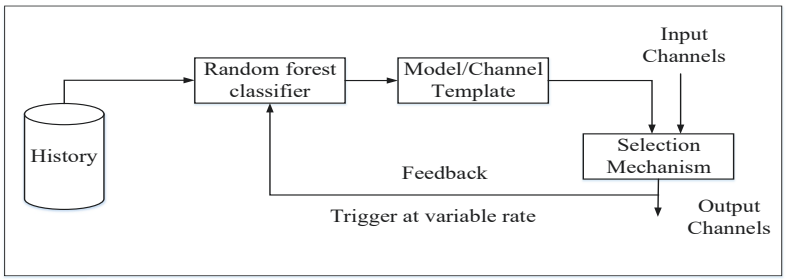


Figure 5.18: ProML Schematic Diagram

- TFR: Transmission Frequency Response.
- RFR: Reception Frequency Response.
- Delay: normalized delay.
- P: Signal transmission power.
- B: Bandwidth.

More features to enhance the model's generality:

- N: Number of available channels.
- S: Sampling interval.

Based on (these parameter values) the following four states will be made using the ProML method:

- Busy: An event when a radio channel is genuinely busy and in a proper usage.
- Jammed
- Idle
- Congested

The simulation model includes the following steps:

- **State Simulation**:

This step creates a test data-set by developing (feature values) in the simulated spectrum for every state or class:

DS = {TFR-data, RFR-data, BW-data, Delay-data} . For every DS element (e.g. TFR-data), the set will include (respective values) for the four states Busy, Idle, Jammed and Congested.

A typical data-set created using this approach shown at table 5.1.

Table 5.1: Typical Data-Set Example

	Features					Classification labels				Sample size
Data-set Type	Transmit Power	Transmit Frequency.	Signal Bandwidth	Available Channels	Delay	Idle	Busy	Congestion	Jammed	
Idle	17 kw	100 MHz	100 MHz	2	2 ms	0	1	0	1	500
Busy	....	....	....	...	..	0	0	0	1	500
Congestion	...	....	....	...	...	0	1	0	0	500
Jammed	...	....	....	...	...	0	1	0	0	500

- **Learning and Validation by Offline State Detection**:

In this step, a classification approach using Random Forest is applied to detect the states of nodes being one of Busy, Idle, Jammed or Congested. The rationale behind using the Random Forests includes:

- Faster than most of the classifiers used for Q-learning.
- Requires relatively small sizes of data to learn.
- Outputs tree hyphen like structures to represent nodes in either Busy, Idle, Jammed or Congested.

Feature values are created based on the reference (Hoang et al., 2017) where minimum and maximum thresholds for every state are defined. As an example power value = 17 kw means it can be a jamming attack as matrix $\begin{bmatrix} aa & bb & . & . & nn \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 3 & 7 & 18 & 21 \end{bmatrix}$ serial. Figure 5.18.

Expected Output is shown at Table 5.2.

Table 5.2: Expected Outcome

Time	Channel.1	Channel.2	Channel.3	Channel.4	.	.	.	.	.	Channel.n
instant	$S_1 = b$	$S_2 = i$	$S_3 = b$							$S_n = j$
T_1										
T_2										
T_3										
T_x										

The classification approach using Random Forest includes:

1. Create a random combination of the data-sets. $D = \{D_i, D_b, D_c, D_j\}$

2. Train the Random Forests algorithm:

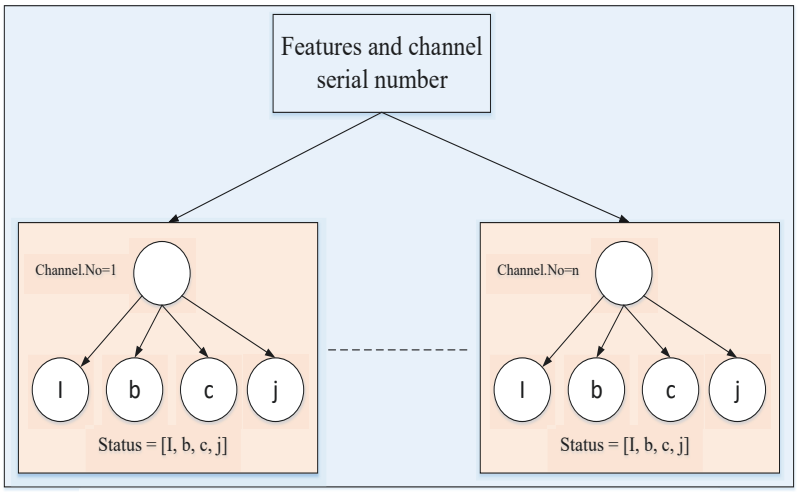


Figure 5.19: Features Channel Selection

3. Generate simulated conditions based on (subject matter expertise) for following state conditions:

- Idle
- In use
- Congested
- Jammed

A typical Random Forest generated tree node as used for this simulation:

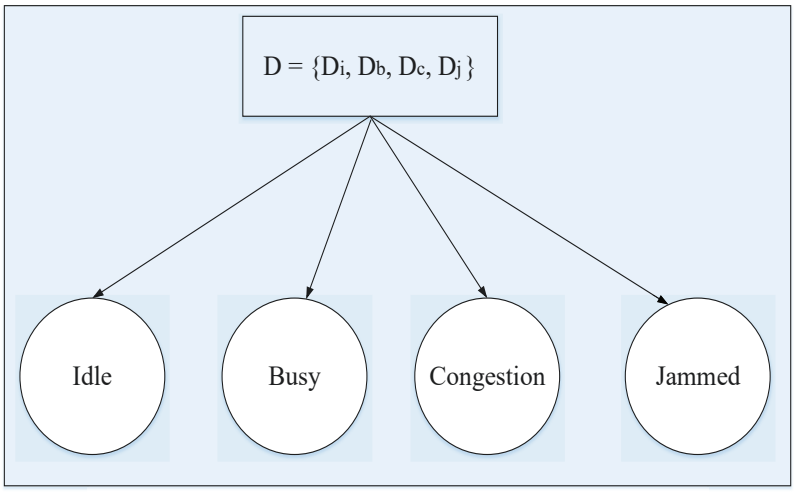


Figure 5.20: Random Forests

4. Testing for current jamming attacks: Once the classification model is trained and validated, the next step is to test with a scenario created by assuming a real-time working Cognitive Radio network. The process involves injecting features (as shown at Section 5.2), at a fixed rate to the classifier, testing its ability to detect channel states. This part of the process is repeated for a number of channels to simulate the scalability of a typical radio (Hoang et al., 2017). Typical tests

include:

- Random jamming: Randomly inducing data-set samples of D_j
- Constant Jamming: Use of very high proportion of D_j
- Deceptive: Using the samples from $D_j$which are closely comparable to D_b
- Reactive Jammer: In this case, idle and busy states are continuously tested and the jamming noise D_j is introduced only when the D_b signals are sensed.

5. Creating Adaptive sampling strategies: At every nth time, a sample of the data-set D_j is tested for possible jamming. In the case of a specific nth time sample showing the presence of jamming then the sampling rate is doubled. The results of sampling the data-set D_j are again tested for jamming which, if it occurs, then further sampling is doubled and so on. Multiplying the sampling rate ceases when a corrective action, utilizing idle channels (bypassing jammed channels) is applied.

6. Verifying the method for scalability (e.g., increasing number of channels from 25 to 100), accuracy and reliability.

This is a supervised learning approach to prepare the security of the

CR network to protect the radio nodes from possible jamming attacks. The detection scheme shown at Figure 5.21 is later enhanced to learn about the radio states, while operating, and then to apply a new structure of node selection.

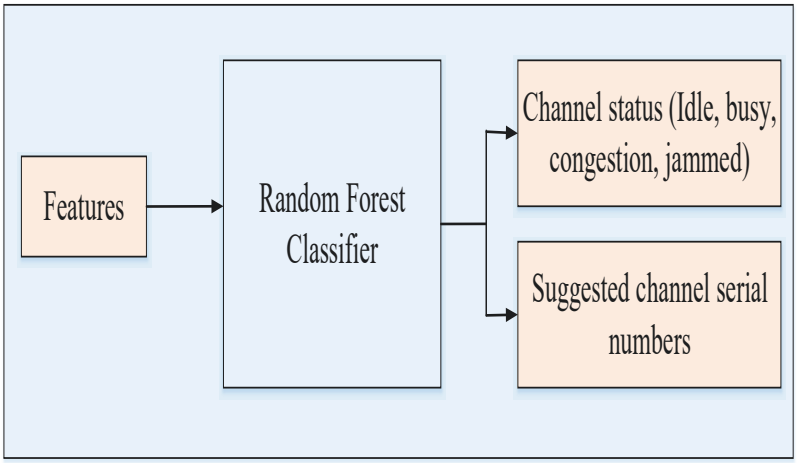
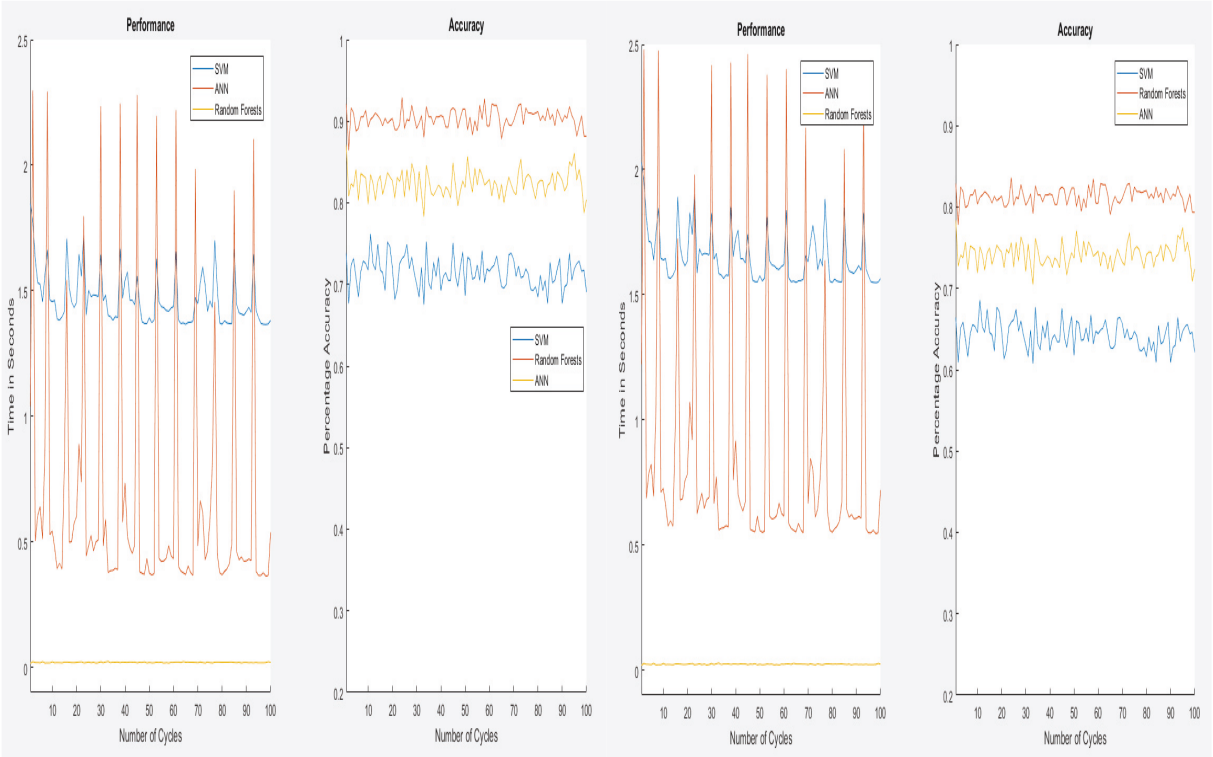


Figure 5.21: Classification Process

5.7.3 Experimental Simulation

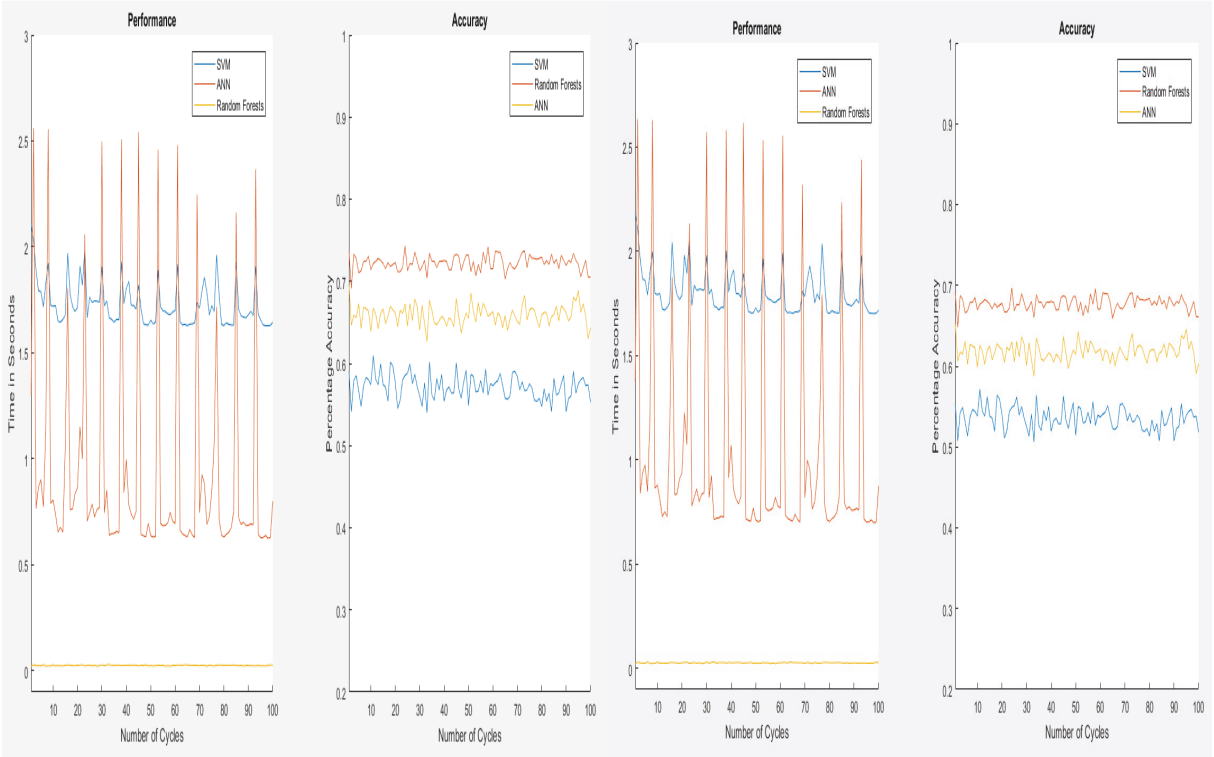
The simulation validates: accuracy, performance and scalability when ProML is applied over the data-set generated during the previous stage. Every 100 simulation cycles, the accuracy and performance comparison is carried out between Random Forests, Feed Forward Perception Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Rationale for using these three classifier is due to their superiority over the other contemporary methods. ANN uses a

feed forward perceptron network, which is one of the fastest methods (Ramesh et al., 2009), whereas a nonlinear SVM kernel is used for this experimental work since the nature of the problem is random and nonlinear. The performance of the two methods are compared with the performance of the ProML method which has Random Forests as its core. The simulation was run for 25, 50, 75 and 100 channels, respectively, selected from a range of channels from 25 to 100. The results are depicted at Figure 5.22 and clearly show the merit of the Random Forests approach (Boulesteix et al., 2012). The range of the channels was considered since we are trying to validate conflicting parameters like accuracy and performance to find advantages of the different classification methods. Thus, scalability at large with accuracy and performance are tested.



(a) **(b)**

Figure 5.22: Average Performance and Accuracy Comparison Between Three most Common Classification Algorithms (Random Forests, Support Vector Machines and Artificial Neural Networks)



(a) **(b)**

Figure 5.23: Average Performance and Accuracy Comparison Between Three most Common Classification Algorithms (Random Forests, Support Vector Machines and Artificial Neural Networks)

The simulation was also conducted to simulate four common types of jamming attacks. For every type of attack considered, the number of channels were 25 and 100 primarily to compare the accuracy of improvement to the nearest accurate traditional swapping method "channel can decide to swap to another channel" (Busch and Malhotra, 2012). The Random Forest algorithm provides accurate results since it is deployed to overcome the dynamism due of jamming attacks (see section 2.10.1). Summary of percentage improvement (of jamming at channels) is shown in the table below. The simulation shows improvement (ProMl Approach) between 2% to 5.1 % over traditional

methods. The constant jamming method is the biggest beneficiary of the proposed methodology (section 5.5) whereas random jamming has a lesser effect on blocking channels. Table 5.4 shows the percentage of jamming at channels.

Table 5.3: Percentage Improvement

Number of channels	Random jamming	Constant jamming	Deceptive jamming	Reactive jamming
25	3%	4.2%	2%	4.1%
50	3.5%	5.1%	2.4%	4.2%
75	2%	4.8%	2.3%	4%
100	2.6%	4.7%	2.2%	3.9%

6 Action Research

The action research describes a broad spectrum of analytical, evaluative, investigative, comparative and bench marking research approaches specifically composed and designed to identify and diagnose issues, challenges, benefits and drawbacks whether related to the specification design, academic, organization or instructional- and support researchers and/or educators in finding viable and effective solutions quickly.

This chapter shows case studies formulated on methodology presented in chapter 5, Section 5.2. The applicability to evaluation algorithm had been tested in Cognitive Radio environment. The framework for the case study is a simulation of remote Cognitive Radio system complete with "attacker" and jamming that work through the scenarios defined for attackers and jamming , in chapter 5, Section 5.2. The idea of Competing Cognitive Radio Network is the system within a communicator node, and jammers attempting to control the admittance to provide a "friendly"environment from the logical inconsistencies of

unfriendly adversaries. In this manner Multi Armed Bandit issues are tested in a variety of scenarios. The purpose for conducting this case study is to examine the proposed solution several times while improving it and then to compare the generated results from each time of testing in order to determine which is the most accurate.

There the framework provides:

- A logical structure for various testing, and
- A perfect Bayesian setting for Thompson Sampling (as described at section 6.4)

The model framework and suppositions around the construction include:

Two theoretical radio systems, Ally and Enemy Competing Cognitive Radio Networks, containing two types of subjective radio node, communicators and jammers.

The Competing Cognitive Radio Networks strives to provide the greatest measure of information by modifying its communication transmissions to the Enemy's jamming activities. The competing Cognitive Radio Networks attempts to cut the Enemy's information throughput by jamming each other's communicator actions. To create a winning media access strategy, they cooperatively optimize anti-jamming and

jamming systems.

Secondly, Mobile ad hoc networks define the networking model for this experiment. The incomplete infrastructural provision for the nodes is presumed. The competing Cognitive Radio Network can implement a centralized regulator model where the node activities are calculated by a particular producer to guarantee an intelligible, network-wide approach. From another side, a distributed regulator model allows every node to calculate its activity. We study both regulator models.

Thirdly, Figure 6.1 demonstrates a periodic multi-channel band for open access. There are N non-covering channels set at the center recurrence f (MHz) with transmission capacity B (Hz). Every time-frequency slot provides a transmission opportunity of interval T msec. We adopt that a node can capture other nodes' transmissions in series. Like identifying ability, but it is not joined to details of any predictable media access regulator mechanisms such as Carrier-Sense Multiple Access.

Fourthly, The Reward System will gauge tasks for a Competing Cognitive Radio Network. A communicator node obtains a reward of B (bits) upon a successful communication, which needs only a single communicator hub communicating in the whole space without being jammed. If there is more than one transmission per unit time, interference will occur, no communicator hubs will receive a reward.

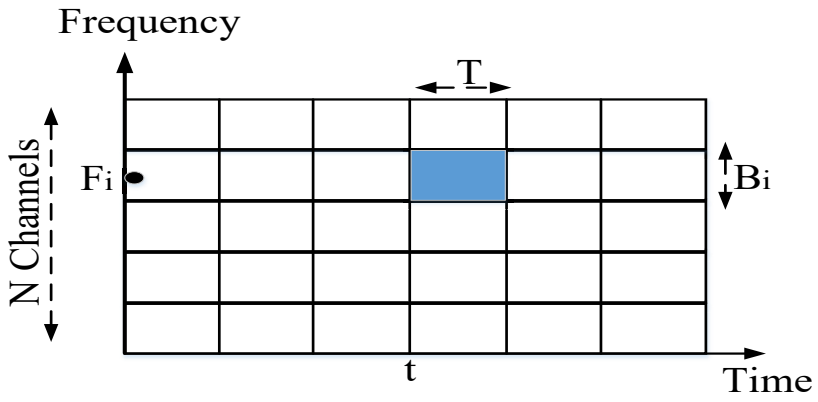


Figure 6.1: Transmission Prospects in Multi-Channel Band Process

Example, Ally jammer obtains B when it jams an Enemy communicator node communicating B-bit value of data. If there were no jamming, the Enemy communicator would have earned B. Also it is likely for Ally jammer to jam Ally communicator mistakenly which is known as sim-jamming.

6.1 The Design for Competing Cognitive Radio Networks

Thompson utilized the Multi-Armed Bandit problem to report the issues in a medicinal preliminary of a clinical trial creating distinctive impacts to patients and other participants (Thompson, 1933). This section demonstrates the Multi-Armed Bandit methodology for Competing Cognitive Radio Networks work with the objective of collecting uncorrupted data (ideal rewards from unknown parameters) of

the channel node relations that required to be learned consecutively.

Multi-Armed Bandit role plays with a card shark fronting N slot machines. The card shark's point is to determine a methodology that misuses the machine to gain maximum rewards.

6.1.1 Multi-armed Bandit Model for Competing Cognitive Radio Network

The Multi-Armed Bandit problem has two forms:

- Centralized control and,
- Distributed control.

Centralized Control Cognitive Radio Network

The multi-armed bandit model for competing for Cognitive Radio networks is a help looks like a direct in the range underneath battle. Communicator hubs and jammers are the contenders that the frameworks allot to run the channels. In the interim, each framework has numerous hubs; our issue is different as a multiplayer Multi-furnished Bandit, which is not at all like the standard one contender Multi-outfitted Bandit communicated by (Lai and Robbins, 1985). Besides, we have two structure contrasts restricted by a unified control element or appropriated control substance. Shows the Competing Cognitive Radio Network with a focal choice maker computing the system wide

arrangement and disseminating all hub exercises. It assumes for the brought together multi-player Multi-Armed Bandit that the choice maker would most likely amass recognizing results from every contender and the precise after-effect of each rival toward a path to make ends toward the end. Figure 6.2 illustrated the centralized control. In this figure we have the central decision maker (head node) compute the all wide network and distribute all action of all nodes, the MRB identifies the decision maker which has to gather sensing data from all nodes (players) in order to make a decision over time.

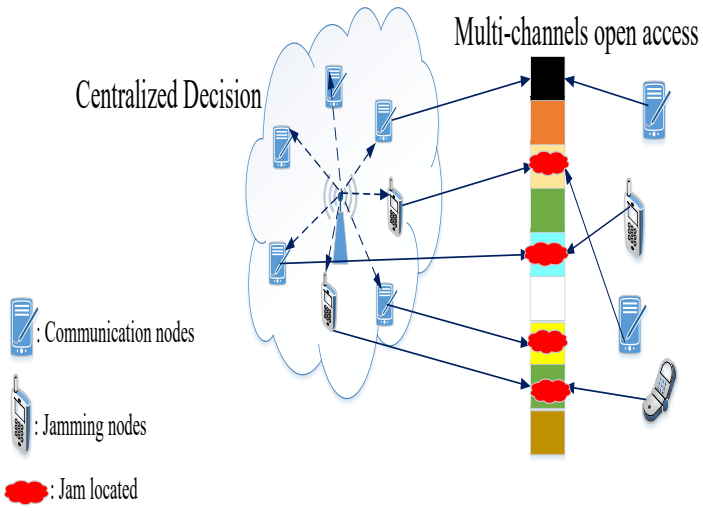


Figure 6.2: Centralized Control Cognitive Radio Network

Distributed Control Cognitive Radio Network

Figure 6.3 depicts the circumstances for "scattered basic leadership". Every hub makes its decision based on information created in top

struggle associated to the centralized multi-player Multi-armed Bandit that needs the proper intro-network communication to gather data and distribute the plan. After each play, the hub (node) observes the outcome, calculates any reward, and keep the player state that can share with others in the network system.

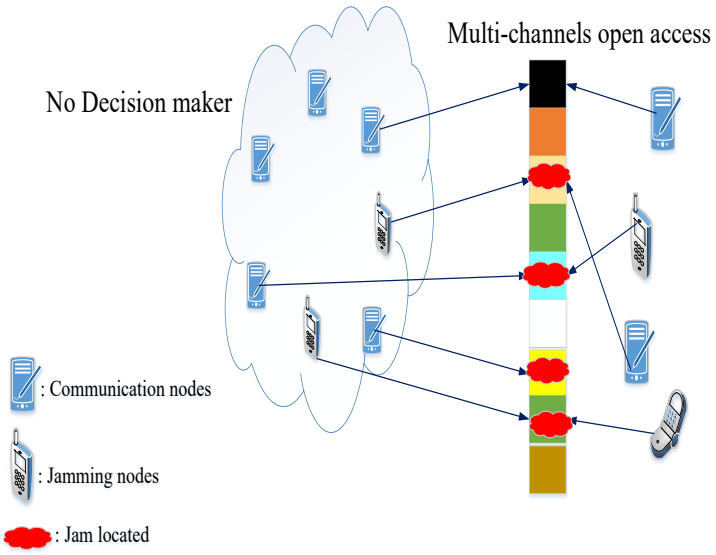


Figure 6.3: Distributed Control Cognitive Radio Network

6.2 Algorithm 1 (Lai and Robbins Algorithm)

Lai and Robbins presented the way of total reward R_i^t, and complete number of gets of T_i^t for channel I, and delineates two hopeful channels to compute Node Maximum Point Estimate C_{MPE} and Round Robin C_{RR}, based on the greatest point gauge paradigm and Round

Robin Selection (RR), correspondingly (Lai and Robbins, 1985). The Kullback-Leibler dissimilarity between the two serves a test measurement to finalise the determination (Gwon et al., 2013). The principle of Lai & Robbins is to study mistreatment picking the maximum point estimate candidate vs. study picking the arbitrary round robin candidate. The state Kullback-Leibler divergence guarantees that picking the maximum point estimate candidate is ideal after enough number of examining tribunals.

Algorithm 6.1 Lia and Robbins

1: **Begin**
2: **While** $t < 1$
3: Access each channel (at least one).
4: Record the cumulative reward for every channel i. The cumulative reward equal the sum of reward populations and total number of accesses for each channel i when jammer equals 1 at time t

$R_i^t = \sum_{j=1}^{t} r_i^j$ and T_i^t

5: **end**
6: **While** $t \geq 1$
7: Compute the hypothetical maximum average reward which equal to the reward population divided by the total number accesses), if gambler action were resulting the best possible outcome each round

$\mu_i = R_i^t / T_i^t \forall_i$

8: Find the maximum $C_{MPE} = i^*$ S. t. $\mu_{i^*} = \max \mu_{i^*}$ Find the maximum point estimate candidate, where the maximum point estimate candidate is equal to the maximum of the hypothetical maximum average rewards,
9: Find robin round

$C_{RR} = (t \ \mathbf{mod} \ N) + 1$

10: If the Kullback-Leibler divergence between the round robin candidate and the maximum point estimate candidate is bigger than the log $(t-1)$, divide by the total number of accesses

$D_{KL}(p_{RR} || p_{MPE}) > \mathbf{log} \ (t-1)/T_{C_{RR}}^t$

11: Access C_{MPE} and observe $r_{C_{MPE}}^t$ Then access the maximum point estimate candidate and observe instantaneous rewards for the maximum point estimate candidate at time t
12: Update $R_{C_{MPE}}^t$ and $T_{C_{MPE}}^t$ Update the cumulative reward for the maximum point estimate candidate at time t and update the total number of accesses for the maximum point estimate candidate at time t.
13: If the Kullback-Leibler divergence between the round robin candidate and the maximum point estimate candidate is NOT bigger than log $(t-1)$ divide by the total number of accesses

$D_{KL}(p_{RR} || p_{MPE}) \leq \mathbf{log} \ (t-1)/T_{C_{RR}}^t$

14: Access C_{RR} and observe $r_{C_{RR}}^t$ Then access the round robin candidate and observe instantaneous rewards for the round robin candidate at time t
15: Update $R_{C_{RR}}^t$ and $T_{C_{RR}}^t$

16: **end**

17 : **end**

The figures below illustrated the simulation of the algorithm Lai and

Robbins, using the Omnet++ simulation.

Figure 6.4 showing Algorithm 1 Simulation initial state, which includes hosts and channels.

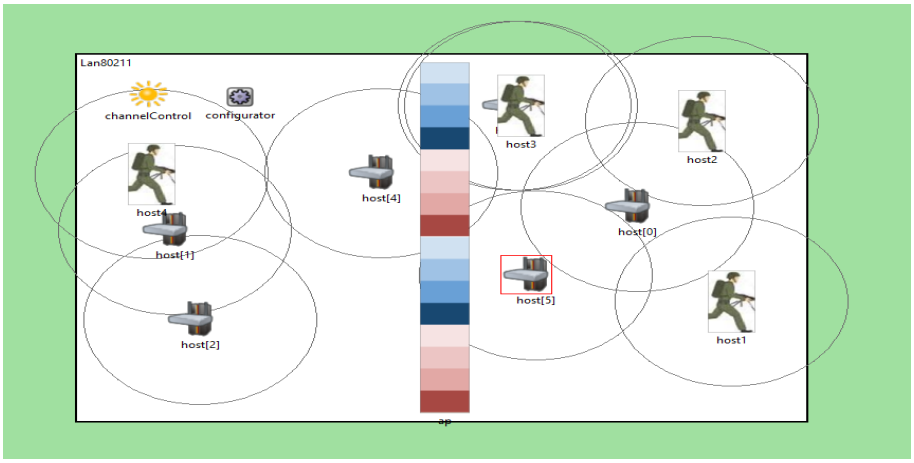


Figure 6.4: Algorithm 1 Running before the Simulation Using OMNET++

Figure 6.5 showing the algorithm simulation running which includes hosts, channels where they have communication between each other.

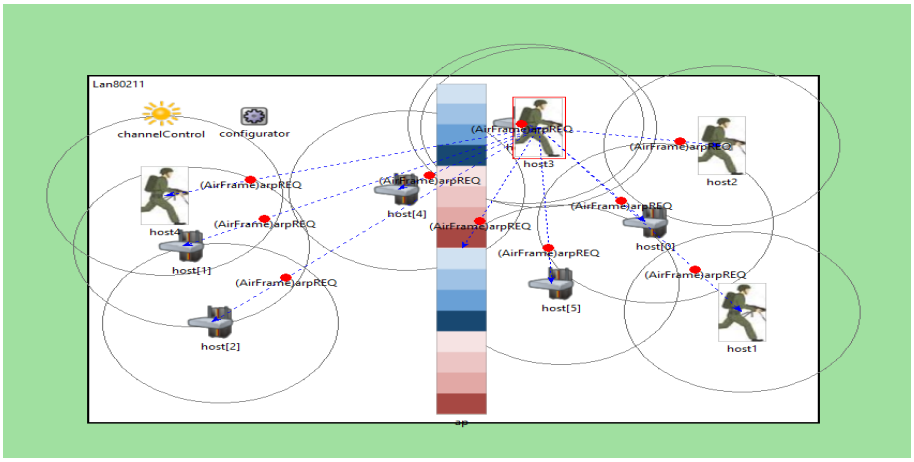


Figure 6.5: Algorithm 1 Simulation Running Using OMNET++

Figure 6.6 showing the result of the simulation of the Access Point after applying Algorithm 1.

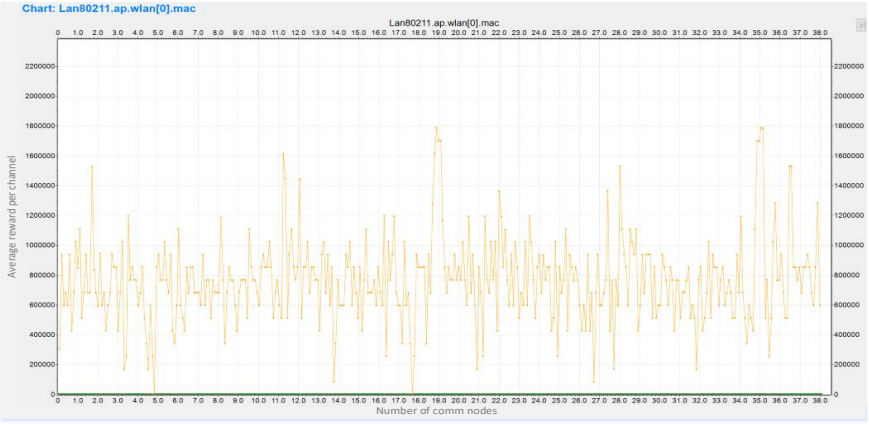


Figure 6.6: Access Point after Applying Algorithm 1Using OMNET++

Figure 6.7 showing the result of packets dropped while doing the simulation for the Access Point after applying Algorithm 1.

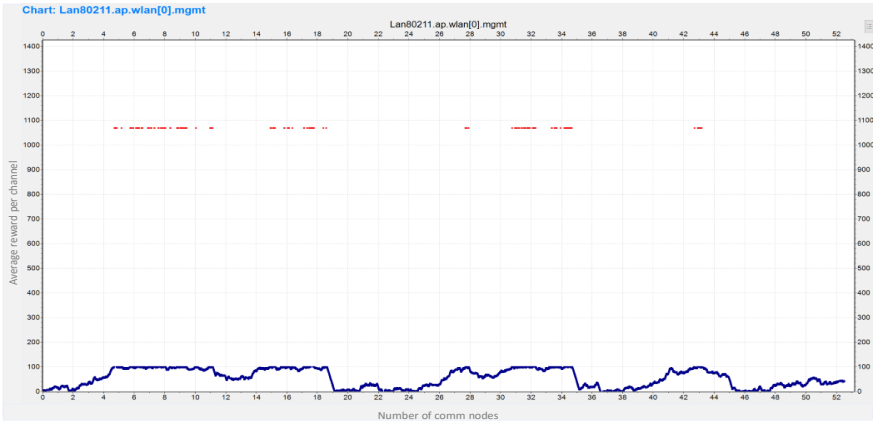


Figure 6.7: Access Point Simulation Showing Packets Dropped Using OMNET++

Figure 6.8 showing the performance for one host in centralized scenario after applying Algorithm 1.

Figure 6.9 showing the performance for four hosts at the same time in centralized scenario after applying Algorithm 1.

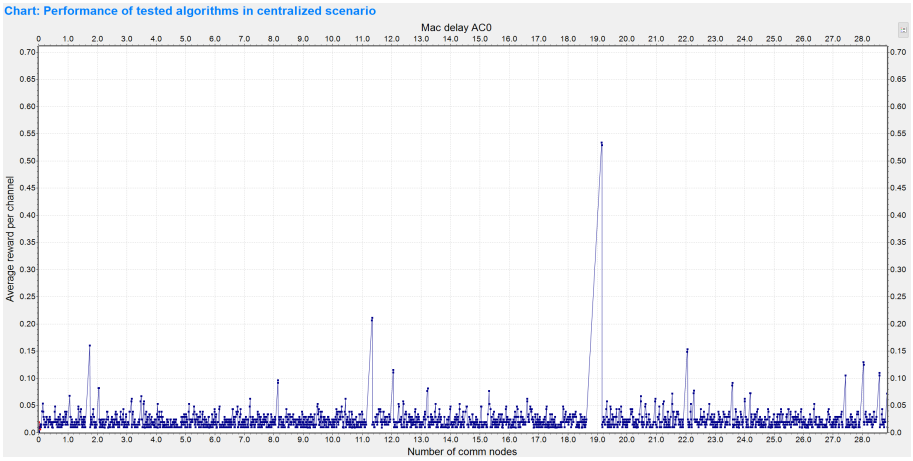


Figure 6.8: Performance in Centralized Scenario for 1 Host Using OMNET++

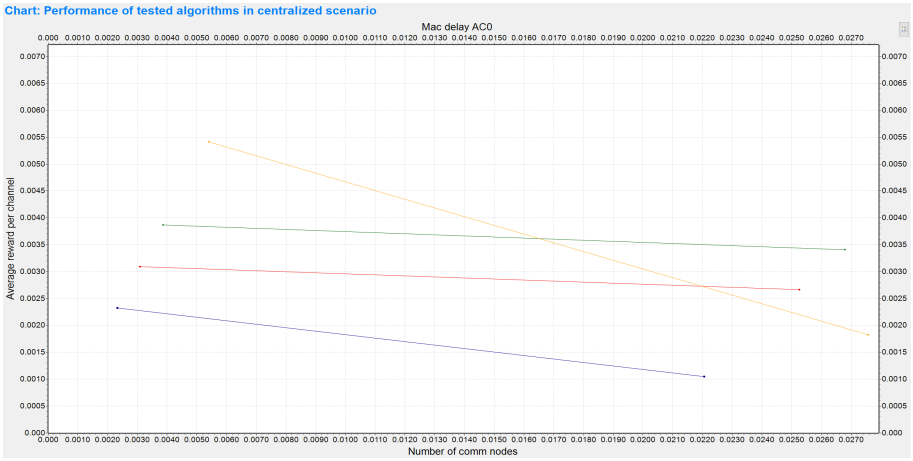


Figure 6.9: Performance in Centralized Scenario for 4 Hosts Using OMNET++

6.3 Algorithm 2 (Upper Confidence Bound Algorithm)

In spite of its algorithmic simplicity, (Lai and Robbins, 1985) drives low to approximating Kullback-Leibler Divergence precisely, which is calculationally hard from sampling. The next class of Multi-Armed Bandit Algorithms utilize indexing as an alternate for Kullback-Leibler Divergence. (Auer et al. 2002) expressed a material structure named

Upper Confidence Bound offered in Algorithm 2.

Algorithm 6.2 Upper Confidence Bound

1: **Begin**
2: **While** $t < 1$
3: Access each channel (at least one), record the cumulative reward for every channel i, where the cumulative reward equals the sum of the reward population and the total number of accesses of each channel i, when the jammer equals 1 at time t.

$R_i^t = \sum_{j=1}^{t} r_i^j$ and T_i^t

4: **end**
5: **While** $t \geq 1$
6: Compute the hypothetical maximum average reward which equals the reward population divided by the total number of accesses.

$\mu_i = R_i^t / T_i^t \forall_i$

N.B. If is a hypothetical maximum average reward if the gambler action resulted in the best possible outcome each cycle.
7: Compute the index equal to the hypothetical maximum average reward plus the square root of $\alpha \log \frac{t}{T_i^t} \forall_i$

$g_i = \mu_i + \sqrt{\alpha \mathbf{log} \frac{t}{T_i^t}} \forall_i$

8: Access Channel equal maximum argmax for index $i^* = arg \max_i g_i$

9: Update the reward population and the total number of accesses.

Update R_{i*}^t and T_{i*}^t

10: **end**

11: **end**

The figures 6.10, and 6.11 illustrate the simulation of the Upper Confidence Bound algorithm

Figure 6.10 shows the algorithm upper confidence bound simulation initial state.

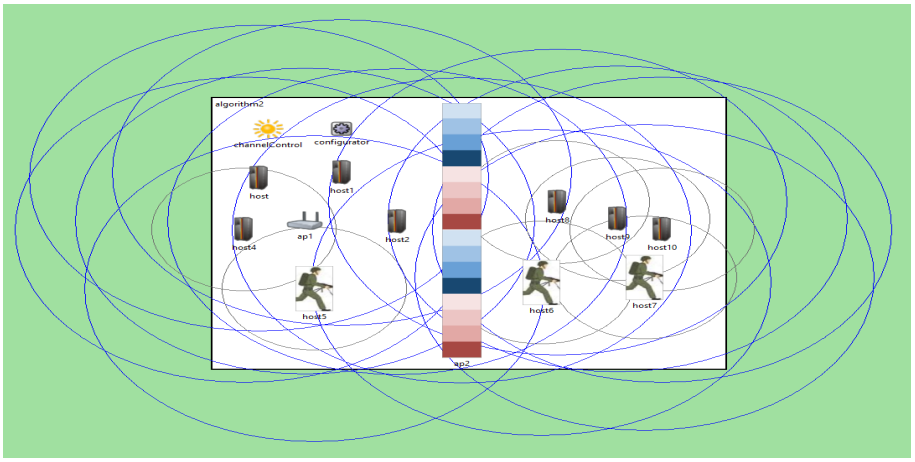


Figure 6.10: Upper Confidence Bound Simulation before Running Using OMNET++

Figure 6.11 shows a simulation of the upper confidence bound algorithm running.

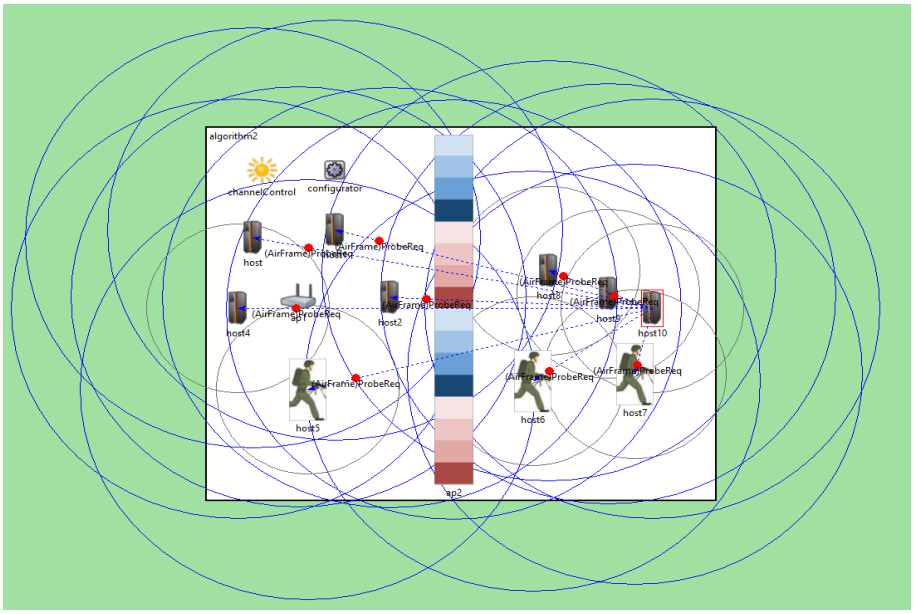


Figure 6.11: Algorithm UCB Simulation Running Using OMNET++

Figure 6.12 shows the result of the simulation on the access point after applying the upper confidence bound algorithm.

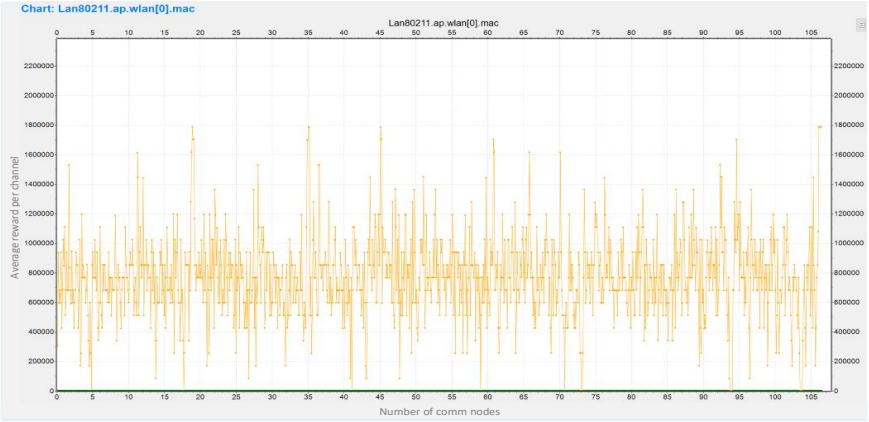


Figure 6.12: Access Point Simulation Using OMNET++

Figure 6.13 shows the result of the packets dropped while running the simulation for the Access Point after applying the upper confidence bound algorithm.

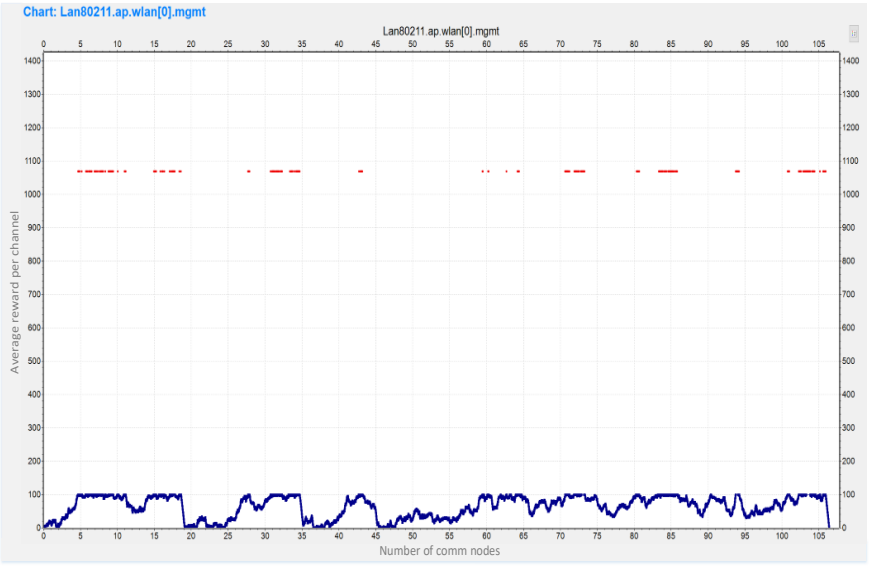


Figure 6.13: Access Point Simulation Showing Packets Dropped Using MONET++

Figure 6.14 shows the performance of four hosts in the same centralized control scenario, after applying the UCB algorithm.

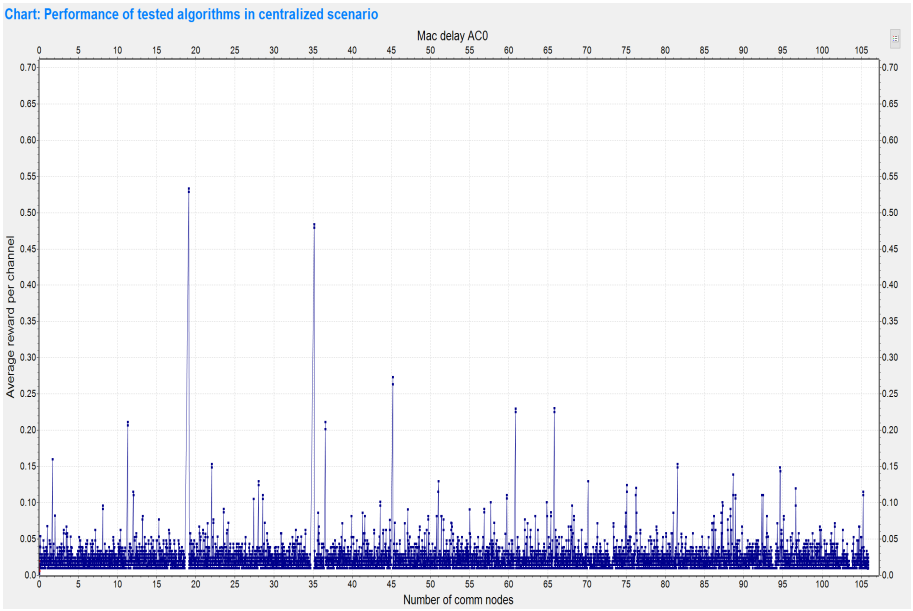


Figure 6.14: Performance in Centralized Scenario for 1 Host Using MONET++

Figure 6.15 shows the performance of four hosts at the same time, in distributed scenario, after applying the UCB Algorithm.

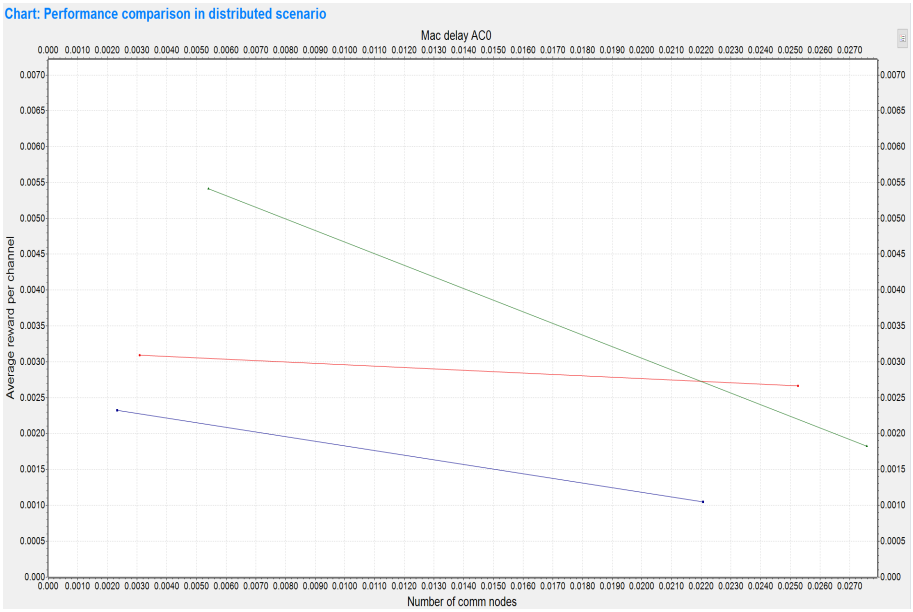


Figure 6.15: Performance in Distributed Scenario for 4 Hosts Using MONET++

Figure 6.16 shows the performance of four hosts at the same time, in

centralized scenario, after applying the UCB Algorithm.

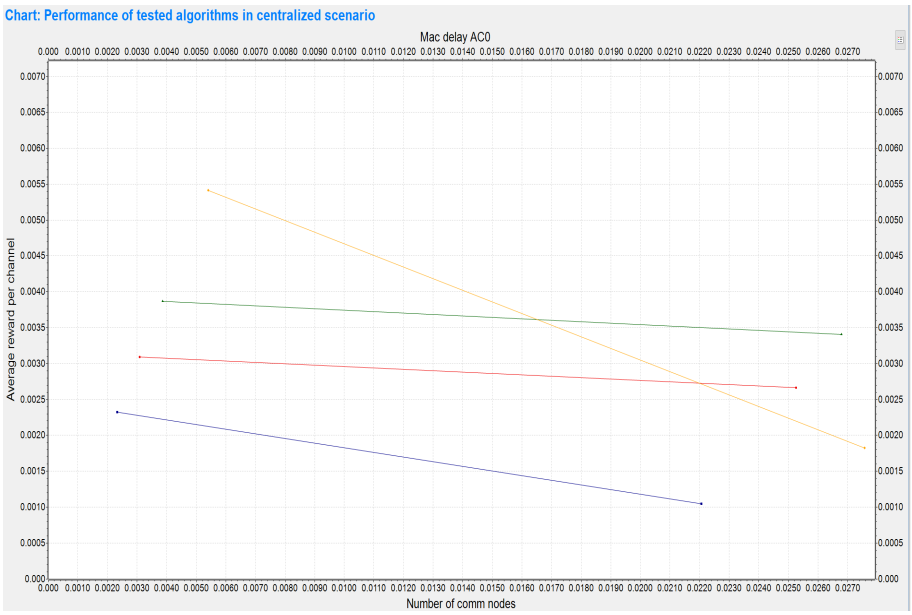


Figure 6.16: Performance in Centralized Scenario for 4 Hosts Using MONET++

6.4 Algorithm 3 (Thompson Sampling Algorithm)

The last level of algorithms practices a chance corresponding method identified as Thompson Sampling that chooses activities conferring to their possibility of presence ideal. It is mostly experimental and has re-stick in the last machine studying info like (Agrawal and Goyal, 2012) which offers the hardest treatment existing today. The full proof of Thompson Sampling on its combining, keeps on being an uncovered issue. It is most famous implicit under a Bayesian setup as in the Thompson sampling Algorithm.

Algorithm 6.3 Thompson Sampling Algorithm

1: **Start**

Require: $d = \{x, a, r\}$ for context x, action a, reward r,
Estimator $p(\theta \mid d) \alpha p(r \mid x, a, \theta) p(\theta)$ parameterized by θ

2: **While** $t \geq 1$

3: Acquire x^t

4: Draw $\theta^t \sim p(\theta)$

5: Choose a^t to access channel equal maximum argmax for index $i^* = arg\max_i \mathbb{E}\left[r_i^t \mid x^t, \theta^t\right]$

6: Observe actual r^t

7: Update $d = d \cup \{x^t, a^t, r^t\}$

8: Update $p(\theta) = p(\theta \mid d)$

9: **end**

10: **end**

The figures below illustrated the simulation of the Thompson Sampling algorithm.

Figure 6.17 shows the result of packets dropped while running the simulation for the access point after applying the Thompson Sampling algorithm.

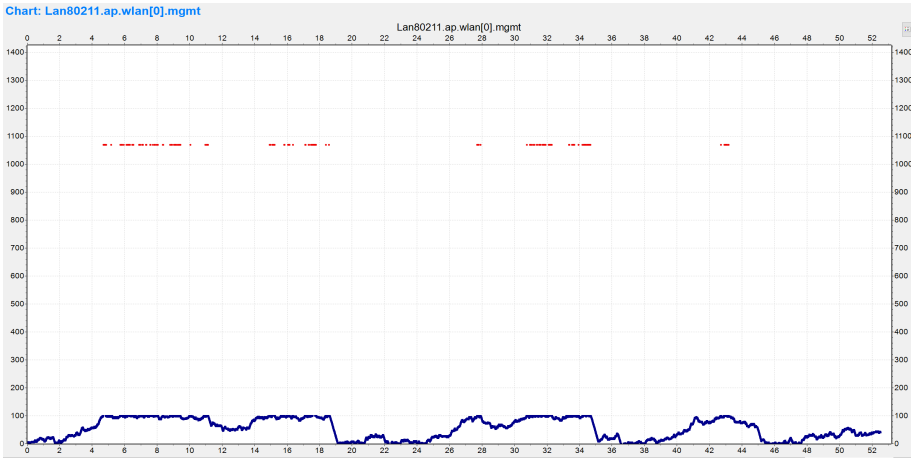


Figure 6.17: Access Point Simulation Showing Packets Dropped Using MONET++

Figure 6.18 shows the performance of four hosts, at the same time, in a distributed scenario after applying Thompson Sampling algorithm.

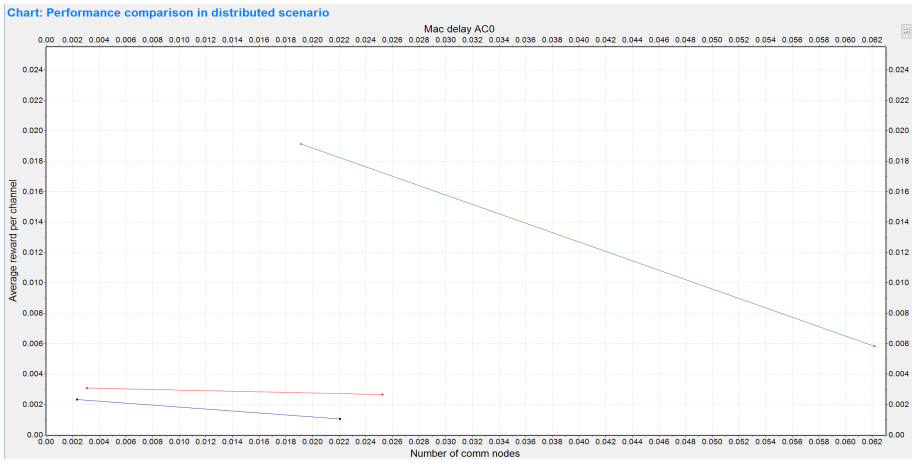


Figure 6.18: Performance in Distributed Scenario for 4 Hosts Using OMNET++

Figure 6.19 showing the the performance for four hosts at the same time in centralized scenario, after applying, the Thompson Sampling algorithm.

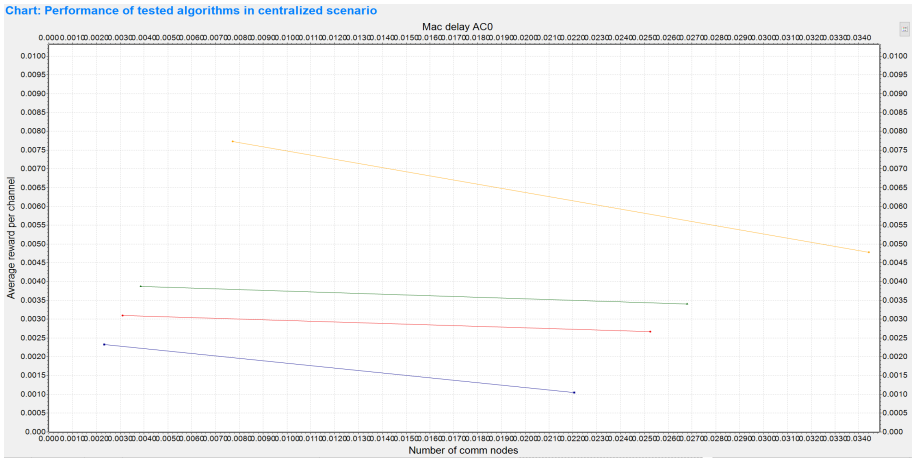


Figure 6.19: Performance in Centralized Scenario for 4 Hosts Using OMNET++

Conclusion:We have defined Competing Cognitive Radio Networks that work under aggressive traditions to struggle for controlling admit-

tance to an exposed field. Our concept of Competing Cognitive Radio Network supports both communications and jamming competences. We have assumed the Multi-armed Bandit framework and carefully inspected perfect results identified to improve an original, optimal media access strategy for Competing Cognitive Radio Network. An optimal Competing Cognitive Radio Network strategy must squeeze randomized algorithms while doing randomization which will lead to weak operation because a strategy wants to develop its knowledge. Our outcomes show that Thompson Sampling shows maximum activity in directing the investigation manipulation trade-off, which is important to make a Multi-armed Bandit-optimal strategy for Competing Cognitive Radio Network. The new future algorithm is more improved than Thompson Sampling, but might steadily surpass the current Multi-armed Bandit algorithms. For our following phase, we plan to study scenarios with the result of activities reliant on unidentified physical settings that direct wireless broadcast performance, more composite reward models, and restriction optimization. Correspondingly, we need to direct our attention to the problems in recognizing errors. Procedure specification and application on wireless software are also need to be done.

The second case study ***Analyzes The Challenges Of Wi-Fi***

Communication And Accessing Free Spectrum Using Cognitive Radio Technology.

Recently, wireless communication technologies such as Wi-Fi communication has received a lot of research attention. As a result, most of the latest devices (laptop, IPad, headset) are equipped with built-in Wi-Fi technology. Wi-Fi frequency spectrum is free, and is used by terminal devices like smartphones, laptop, tablets computers, to remote sensors, actuators televisions and many more. Wi-Fi technology has access to the 2.4GHz and 5GHz frequency bands (Dhivya and Ramaswami, 2017). The 2.4 GHz band offers a higher range than the 5GHz frequency band (Chao et al., 2015). Because the higher the wireless frequency, the higher is its bandwidth. Therefore, 5 GHz frequency spectrum offers higher bandwidth than 2.4 GHz frequency (Lee et al., 2015). Most of severances use the 2.4GHz frequency because in Wi-Fi communications, range is a major factor.

Due to the wide applications of 2.4GHz frequency band, it is jammed more by malicious and non-malicious attackers and it is becoming more challenging to use it. 2.4 GHz Wi-Fi communication suffers from three main challenges, including channel interference, channel congestion and network attack or actual malicious jamming (Chao et al., 2015). To address these challenges, a CR model is proposed

by this study. The 2.4 GHz Wi-Fi signal is analyzed with tools (as describe in section 6.2) for different scenarios using different tools and the possible solutions are discussed and analyzed.

6.5 Main Challenges of WiFi Communication

The 2.4 GHz Wi-Fi suffers from three main challenges, including: co-channel interference, channel congestion and network attack or jamming.

6.5.1 Channel Interference

The 2.4GHz (2400-2500) frequency band is a "free spectrum" (Tursunova et al., 2010). No permission is required from Telecommunication Authorities to use this band. Due to its free access, the number of users utilizing this frequency band is increasing significantly. In 2.4GHz (2400-2500) band, multiple technologies and devices coexist. It not necessary that they use 802.11 Wi-Fi technologies. Some examples of these devices include Bluetooth, ZigBee, Mobile phone, Z-wave, wireless baby monitors and other allied devices (Ozdemir, 2009).

Generally coexisting technologies do not cause interference to each other, because of their own transmission protocols and mechanisms. However, there are chances of interference when all the devices occupy the medium using exactly the same frequency band. While communicating or transmitting data, some devices are more aggressive

than others and prevent communicating by generating interference and noise, resulting in radio-electric spectrum saturation (Ishizu and Harada, 2009).

6.5.2 Channel Congestion

Generally there are 13 available channels in the 2.4GHz spectrum, in most countries. Among these 13, channels 1, 6 and 11 do not overlap with each other (Ozdemir, 2009). Wi-Fi (IEEE 802.11) technology uses the 2.4-2.5GHz spectrum. In Australia, this 100MHz spectrum is divided into 13 channels with 5MHz space between channels (Ishizu and Harada, 2009). Channel 1 starts at 2400 MHz and is centered at 2412 MHz. Channel 2 starts at 2417 MHz, keeping a 5MHz space between the two channels. The width of each channel is 22 MHz which allows only three non-overlapping channels. In the 2.4 GHz Wi-Fi frequency band the non-overlapping channels are 6 and 11 (Jovicic and Viswanath, 2009). In Wi-Fi, the signal is operated in a half-duplex mode i.e. at any given time, only one device can transmit and communicate on a channel for a device to communicate and transmit, it needs to wait for its turn. When two or more than two devices try to transmit simultaneously, they interrupt each other, since only one device can transmit at a time in a channel (Nekovee, 2010). Therefore, it is necessary to limit the number of devices that can connect to any

channel, so that all the devices can communicate and transmit data without disruption.

Presently, the Wi-Fi technology does not support advanced channel allocation, because channels 1, 6 and 11 are not overlapped. When the devices are turned on, most of those devices select one of these three channels automatically (Ishizu and Harada, 2009). As an outcome, three channels become congested while other channels are not being utilized. Figure 6.20 illustrates Wi-Fi channel allocation in the 204 GHz band.

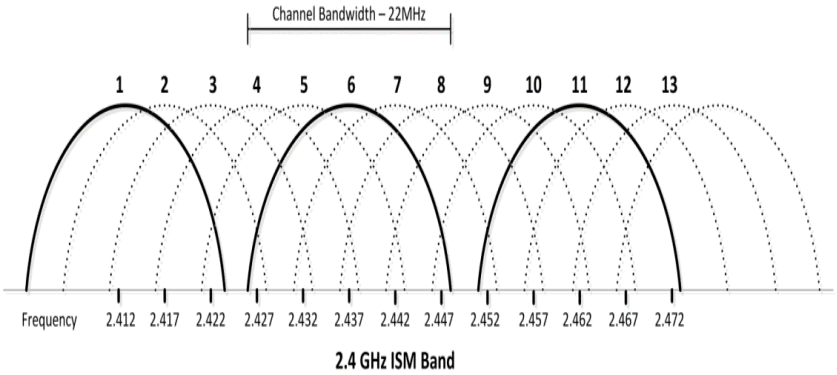


Figure 6.20: Wi-Fi Channel Allocation in 204 GHz, adapted from (Miucic, 2018)

6.5.3 Jamming the Network

Network attacks prevent devices from communicating and transmitting data. If any device uses more than 70% of the bandwidth thus constantly occupying the channel (Shellhammer et al., 2009), other

devices will discover that the channel is busy and will not be able to communicate and transmit data. 802.11 Wi-Fi technology has only one device communicating at any instance of time in a channel (Huang et al., 2010). Jamming can be performed in both intentional (malicious) and unintentional (non-malicious) ways. In jamming, if any device wishes to jam any network, it can send back-to-back packets without following the Inter Frame spacing rule. Consequently, when other devices are verified by the medium to transfer data, they find it occupied and are keep waiting for their turn to communicate. Another kind of attack can be done by transmitting a large stream of 1's and 0's (bit stream). This is known as the Queensland attack (Sridhara et al., 2008).

6.6 Wi-Fi Analysis Tools

Wi-Fi performance and the challenges of using Wi-Fi is analyzed by professional tools. The main tools include Wi-Fi Analyzers, Acrylic Heat Maps and Channelizers.

- **Wi-Fi Analyser**

Analyzes the Wi-Fi spectrum and channels. It shows how many devices are connected to a channel including the Received Signal

Strength Indication (RSSI) of every Wi-Fi signal, Basic Service Set Identifiers (BSSID), and Service Set IDs (SSID) of Wi-Fi signals can also be seen using this tool.

- **Acrylic Wi-fi Heat Maps**

Acrylic Wi-Fi heat maps check Wi-Fi network performance, network cover, and channels usage, including the capture of all data traffic.

- **Chanalyzer Essential**

Chanalyzer Essential can capture raw Wi-Fi spectrum data, be able to determine jamming devices and provide Wi-Fi network covering. It also captures noise or interference in the Wi-Fi band.

6.7 Analysis of Wi-Fi Challenges With The Tools

- **Channel Congestion**

Figure 6.21 shows a number of devices connected to the CR spectrum channels. The graph captured by Wi-Fi analyzer, where channels 1, 6 and 11 utilized by 11, 4 and 9 devices sequentially while other channels are hardly utilized . Among Channels 1, 6 and 11, channel 1

is extremely utilized. Therefore, users of channel 1 might suffer from late connections, effectively slowing down traffic through the band.

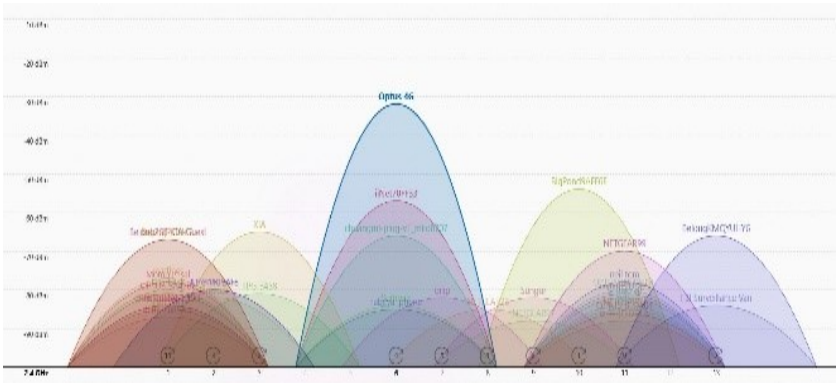


Figure 6.21: Wi-Fi Channels using Analysis WiFi Tool

Figure 6.22 shows all devices are utilizing channel 1, 6 and 11 while other channels are not correctly utilized. There is a great possibility of overlapping access points with each other, and the users will suffer from late connection.

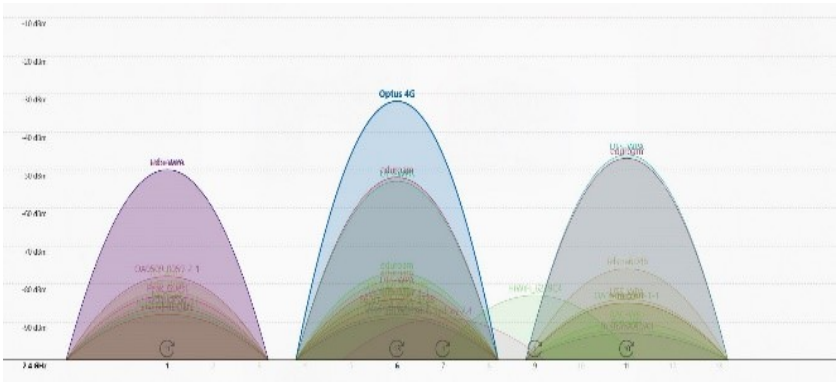


Figure 6.22: Wi-Fi Channels using Analysis WiFi Tool

- **Channel Interference**

Figure 6.23 shows the Channel interference or noise in channel, this

noise cause late connection to user. Using the Chanalyzer and Acrylic Wi-Fi HeatMap tools to capture the channel interference.

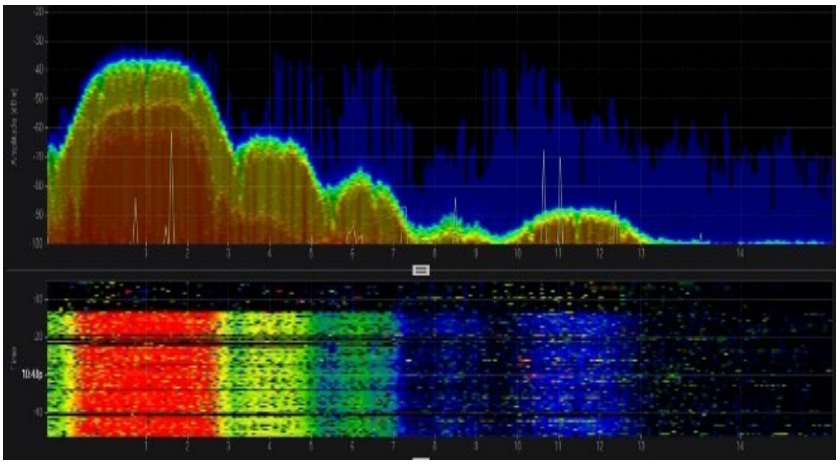


Figure 6.23: Noises on Channel 1 using the Chanalyzer and Acrylic Tools

.

This shows the demonstration of channels, and also shows the noise. This noise is coming from other devices that use the 2.4 GHz spectrum. The red color shows a regular transmission. Channel 1 is 60% usage. So, this interference will cause late connection.

Figure 6.24 showing the noises on channel 12 the same as channel 1, and channel 12 high usage and small interference causes low and late connection for the users.

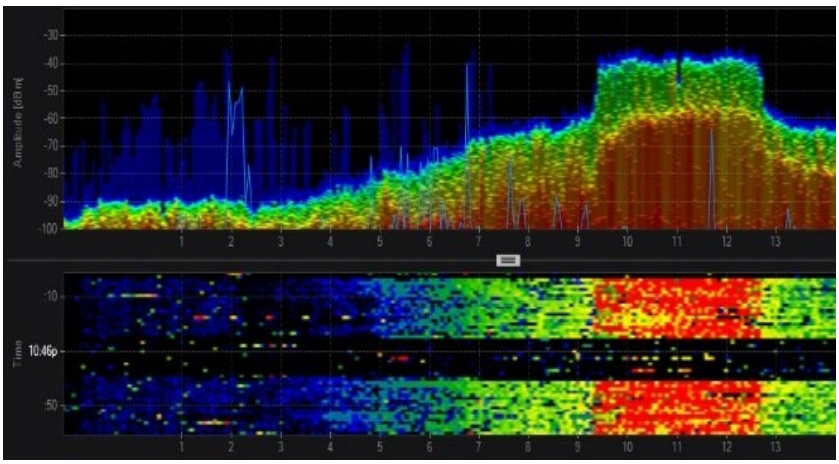


Figure 6.24: Noise on Channel 12 using the Chanalyzer and Acrylic Tools

Figure 6.25 shows the signal cover of a shopping mall area. Using the Acrylic Wi-Fi heatmaps tool , red areas describe the High Relative Received Signal Strength (HRRSS), green areas describe low HRRSS, and also describe the interference. The users in the green area are expected to have low connection with the internet.

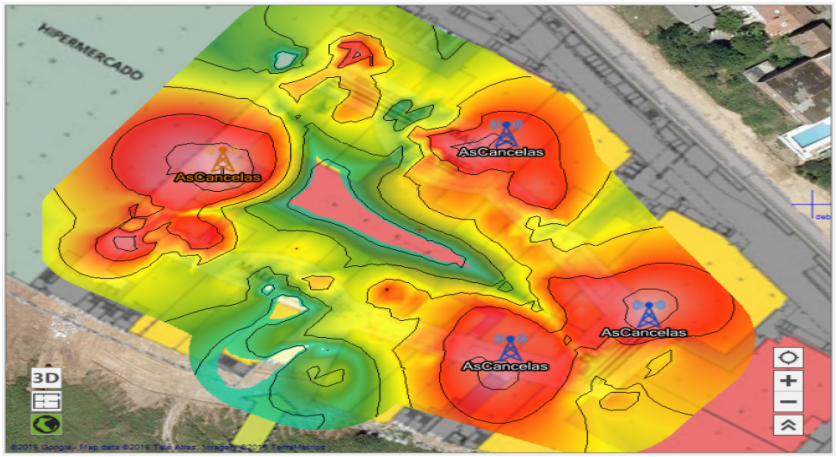


Figure 6.25: Heatmap of RSSI using the Acrylic Wi-Fi Heatmaps Tool

Figure 6.26 shows the 3D graph for the capacity of interference can be shown; the red describes the high signal strength, the blue area

describes the interference.

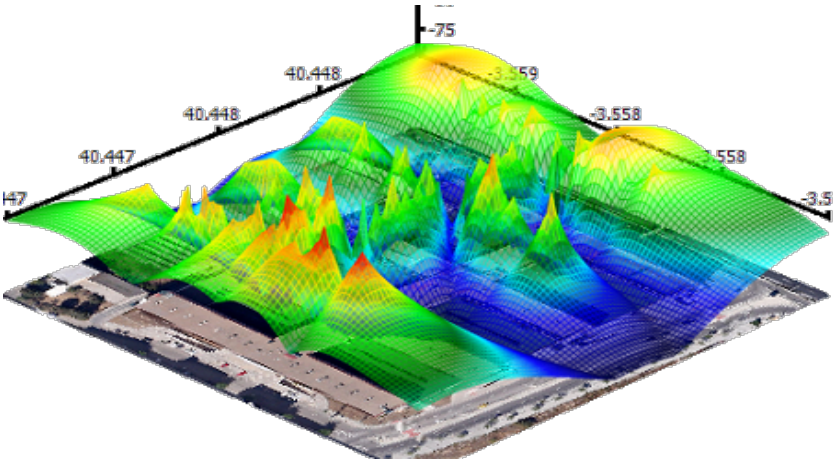


Figure 6.26: Heat Map 3D using the Acrylic Wi-Fi Heatmaps Tool

Figure 6.27 showing the Signal-to-Noise Ratio (SNR) is the useful parameter for measuring communication state because it takes into account signal strength and noise in the wireless medium, it ranges from (0 worst) to (100 best). 60 or above is taken as good value. The red area and green area are describing the highest SNR, the blue/black areas describe low SNR.

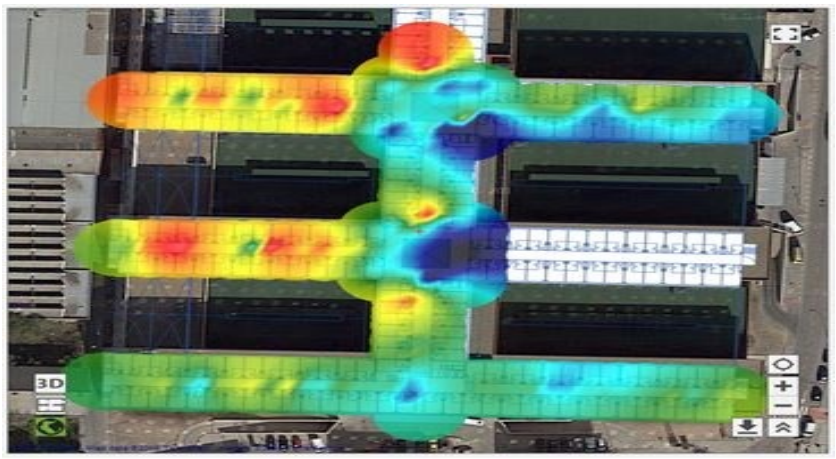


Figure 6.27: Heat Map of SNR using the Acrylic Wi-Fi Heatmaps Tool

Figure 6.28 shows 3 dimensions (3D) map of Signal-to-Noise Ratio

(SNR), and describes the highest and lowest SNR areas.

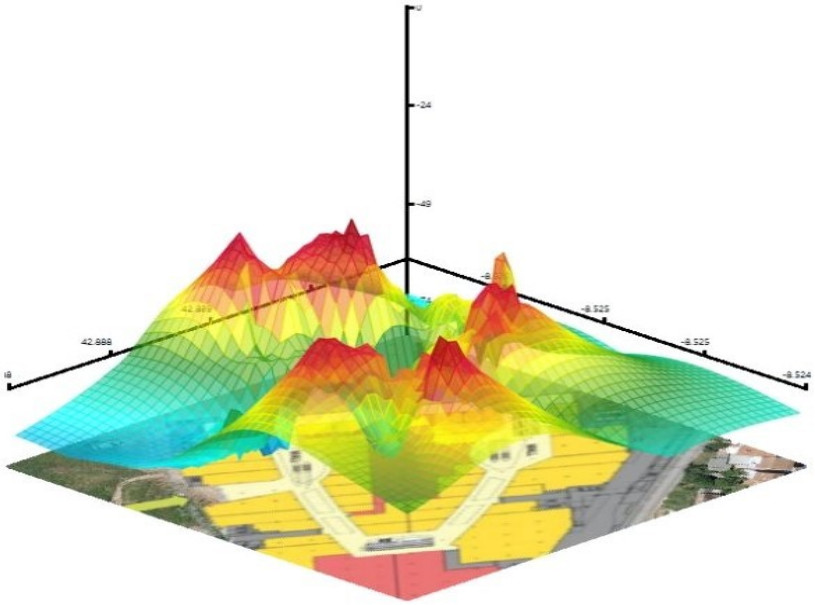


Figure 6.28: Heat map of SNR in 3D Using the Acrylic Wi-Fi Heatmaps Tool

- **Channel Jamming**

Figure 6.29 shows many devices on the first channel, transmitting continuously indicating that the channel is more than 60%, utilized; as an outcome, other devices must wait for their turn to use the channel.

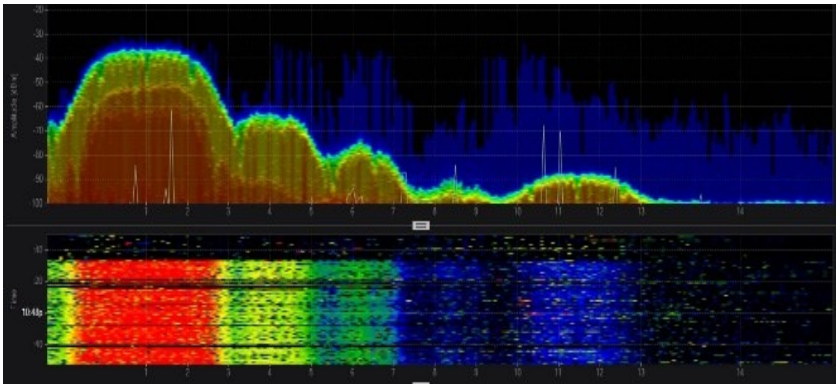


Figure 6.29: Jamming at Channel One using the Chanalyzer Tool

Figure 6.30 shows all the channels being jammed by security cameras. It shows all channels, red color, and all devices find that the channels are busy, and devices will not transfer information.

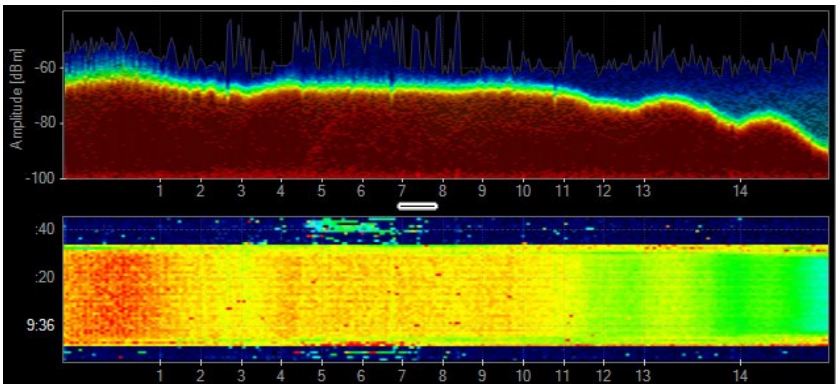


Figure 6.30: Jamming in all Channels using the Chanalyzer Tool

- **Problem Solution**

Cognitive Radio can be applied for Wi-Fi communications to resolve the challenges previously discussed. Channel interference, channel congestion and jamming issues may be resolved by:

- **Management channel interference**

When there are noises and interference in Wi-Fi frequency, Cognitive Radio can be employed to switch between frequency bands. For example, when interference occurs in the 2.4GHz frequency spectrum, Cognitive Radio can switch to the 5GHz frequency spectrum. Con-

versely, Cognitive Radio can automatically switch the wireless transmit power. Therefore, when transmission low bandwidth CR is required, transmission capacity can be reduced, and when required to transmit high bandwidth, CR can increase transmission capacity. Applying this method, CR can mitigate noise or other interferences by raising the transmission power.

- **Channel Congestion**

Popular Wi-Fi technology does not promote effective resource management. Cognitive Radio can verify the spectrum prior to building a connection. The outcome will ensure that all the channels are used, roughly equally, minimizing channel congestion.

- **Jamming Network**

Cognitive Radio regularly reviews differences in its environment. In case of sudden jamming of Wi-Fi channels, CR may resolve the problem directly by switching to the 5GHz spectrum or switching down to available channels.

The challenges of connecting to Wi-Fi are at the moment slow, and users are not making the most of their internet connections. The appli-

cation of CR in the Wi-Fi network can ensure smooth communication and increased usage of the frequency spectrum. Access to licensed frequencies such as white TV space, military white space, may enhance the network, in the future.

7 Conclusion and Future work

This research work has shown that Cognitive Radio technology is able to enhance wireless access based networks, by providing the means of spectrum sensing and its analysis, for optimal performance. However, Cognitive Radio Networks have security issues that need to be addressed. This thesis focuses on one of the most challenging and persisting issues, such as the Jamming Attack problem. This study explored, experimented and validated the original methodology that utilized the MAB problem based strategies for mitigating jamming attacks. These strategies were designed, implemented and validated using 3 different test environments such as: Python (Stewart, 2017) programming environment, the OMNET++ simulation tool (Varga, 2010) and the ProML Matlab environment (Trefethen, 2000). Python programming environment allowed for testing software implementations of the MAB problem based algorithms, whilst the OMNET++ simulation that allowed for a creation and testing of realistic jamming attacks scenarios. In ProML environment it was possible to validate the Machine Learning (ML) based approach for mitigating jamming

attacks. The approach was designed, modelled and coded as the Matlab based algorithm. The solution was tested and compared against the traditional swapping methods for scalability of 100 channels. Results indicate that the ML approach, is more suitable for large networks (near to or greater than 100 channels) rather than for smaller ones - the smaller scalability model is comparable with other swapping methods.

7.1 Outline of Contributions and Main Findings

Major contributions and research outcomes include:

- *A general overview of Cognitive Radios security issues.* There are many threats to Cognitive Radio systems including traditional threats such as: eavesdropping, impersonation, selective forwarding, sinkhole, wormhole and hello flood attacks. Others include: Trend Threats such as software attacks, hardware attacks, and (OSI communication model) Layer Threats such as network layer attacks, transport layer attacks, link layer attacks. In addition Cognitive Radio is based on defined software radio architecture and inherits the security problems associated with that technology. Moreover, Cognitive Radio is sensitive to threats associated

with legacy radio systems, which arose because of its open nature.

- *A theoretical approach to the game theory.* Strategies of the Multi-Armed Bandit problem were employed to formalize intelligent jamming and anti-jamming.

- As current Cognitive Radio technologies often introduce loopholes which enable Jamming Attacks, *variations of the Multi-Armed Bandit problem were deployed to counter Jamming Attacks*, In this work several other inventive methodologies introduce the experimental work which is the main contribution of this research.

- *A historical data-set* including parameters applicable to the type of jamming attacks employed by the research (see above). This data set can be used to predict the impacts of undiscovered jamming. Using this data-set, it was shown that ProML improved channel efficiency during "constant" and "reactive" jamming attacks.

Publications

Journal Papers

- Shaher Suleman Slehat. and Chaczko, Z., 2016, "Mitigating Nat Holes Vulnerability in Teredo Clients". Journal of Networks, 10(9), pp.521-529. (A Journal- ERA Indexed).

- Zenon Chaczko, Shaher Suleman Slehat, Alex Salmon, 2016, "Application of Predictive Analytics in Telecommunications Project Management". Journal of Networks, 10(10), pp.551-566. (A Journal- ERA Indexed).

Conference Paper

- Z. Chaczko, Shaher Suleman Slehat; T. Shnoudi "Game-Theory Based Cognitive Radio Policies for Jamming and Anti-Jamming in the IoT" conference 2018 The 12th International Symposium on Medical Information and communication Technology, UTS conference

- Shaher Suleman Slehat and Z. Chaczko and A. Kale, 2015, "Securing Teredo Client from Nat Holes Vulnerability" Computer Aided System Engineering (APCASE), 2015 Asia-Pacific Conference on, IEEE computer Society, pp 366-369, July 2015, Ecuador.

- R. Wazirali; Shaher Suleman Slehat; Z. Chaczko; G. Borowik; L. Carrión, 2015 " Objective Quality Metrics in Correlation with Subjective Quality Metrics for Steganography" Computer Aided System Engineering (APCASE), 2015 Asia-Pacific Conference on, IEEE computer Society, pp 238 - 245, July 2015, Ecuador.

- A. Kale; Z. Chaczko; Shaher Suleman Slehat, 2015, " HyMuDS: A Hybrid Multimodal Data Acquisition System" Computer Aided System Engineering (APCASE), 2015 Asia-Pacific Conference on, IEEE computer Society, pp 107 - 112, July 2015, Ecuador.

7.2 Future Work

Research performed for this thesis may be extended as follows:

1. *Security issues associated with self-reconfigurable entities that depend on strategies for the Multi-Armed Bandits problem.* The research has shown that an entity may be "fooled" into incorrect activities. The Cognitive Radio network may make false assumptions, due to faults, and probability of software defects in the learning system. Future work should focus on discovering ways to permit the simulation methodologies to restore their "reasoning" processes by learning from any human intervention.
2. The *Multi-Armed Bandit and ProML approaches are to be further tested and enhanced* by more real-time, real life based case studies.
3. *Improving the Multi-Armed Bandit and ProML approaches* in instances of deceptive jamming

Bibliography

Rajeev Agrawal. Sample mean based index policies by o (log n) regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27(4):1054–1078, 1995.

Shipra Agrawal and Navin Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *COLT*, pages 39–1, 2012.

Rehan Ahmed. *Detection of vacant frequency bands in Cognitive Radio*. PhD thesis, Blekinge Institute of Technology, 2010.

Ozgur B Akan, Osman B Karli, and Ozgur Ergul. Cognitive radio sensor networks. *Network, IEEE*, 23(4):34–40, 2009.

Ian F Akyildiz, Won-Yeol Lee, Mehmet C Vuran, and Shantidev Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer networks*, 50(13):2127–2159, 2006.

Ian F Akyildiz, Won-Yeol Lee, Mehmet C Vuran, and Shantidev Mohanty. A survey on spectrum management in cognitive radio networks. *IEEE Communications magazine*, 46(4), 2008.

Ian F Akyildiz, Won-Yeol Lee, and Kaushik R Chowdhury. Crahns: Cognitive radio ad hoc networks. *AD hoc networks*, 7(5):810–836, 2009.

Eitan Altman, Konstantin Avrachenkov, and Andrey Garnaev. A jamming game in wireless networks with transmission cost. In *Network Control and Optimization*, pages 1–12. Springer, 2007.

SaiDhiraj Amuru and R Michael Buehrer. Optimal jamming strategies in digital communicationsâĂŤimpact of modulation. In *2014 IEEE Global Communications Conference*, pages 1619–1624. IEEE, 2014.

S Anand, Z Jin, and KP Subbalakshmi. An analytical model for primary user emulation attacks in cognitive radio networks. In *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*, pages 1–6. IEEE, 2008.

Venkatachalam Anantharam, Pravin Varaiya, and Jean Walrand. Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays-part i: Iid rewards. *Automatic Control, IEEE Transactions on*, 32(11):968–976, 1987.

Kenneth J Arrow, David Blackwell, and Meyer A Girshick. Bayes and minimax solutions of sequential decision problems. *Econometrica, Journal of the Econometric Society*, pages 213–244, 1949.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

Paramvir Bahl, Ranveer Chandra, and John Dunagan. Ssch: slotted seeded channel hopping for capacity improvement in ieee 802.11 ad-hoc wireless networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 216–230. ACM, 2004.

Behnam Bahrak, Jung-Min Park, and Hao Wu. Ontology-based spectrum access policies for policy-based cognitive radios. In *2012 IEEE International Symposium on Dynamic Spectrum Access Networks*, pages 489–500. IEEE, 2012.

Alfredo Banos et al. On pseudo-games. *The Annals of Mathematical Statistics*, 39 (6):1932–1945, 1968.

John Bellardo and Stefan Savage. Denial of service attacks: Real vulnerabilities and practical solutions in the proceedings of the 12th usenix security symposium washington. *DC, USA Aug*, pages 4–8, 2003.

Richard Bellman. A problem in the sequential design of experiments. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 16(3/4):221–229, 1956.

Donald A Berry and Bert Fristedt. *Bandit problems: sequential allocation of experiments (Monographs on statistics and applied probability)*. Springer, 1985.

Shameek Bhattacharjee, Shamik Sengupta, and Mainak Chatterjee. Vulnerabilities in cognitive radio networks: A survey. *Computer Communications*, 36(13):1387–1398, 2013.

Kaigui Bian and Jung-Min Park. Mac-layer misbehaviors in multi-hop cognitive radio networks. In *US-Korea Conference on Science, Technology, and Entrepreneurship (UKC2006)*, pages 65–73, 2006.

Anne-Laure Boulesteix, Silke Janitza, Jochen Kruppa, and Inke R König. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6):493–507, 2012.

Richard Neil Braithwaite. Rf module for wireless unit configured for self-interference cancellation, July 4 2017. US Patent 9,698,861.

Robert W Brodersen, Adam Wolisz, Danijela Cabric, Shridhar Mubaraq Mishra, and Daniel Willkomm. Corvus: a cognitive radio approach for usage of virtual unlicensed spectrum. *Berkeley Wireless Research Center (BWRC) White paper*, 2004.

Sbastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Sébastien Bubeck. *Bandits games and clustering foundations*. PhD thesis, Université des Sciences et Technologie de Lille-Lille I, 2010.

Niv Buchbinder, Liane Lewin-Eytan, Ishai Menache, Joseph Naor, and Ariel Orda. Dynamic power allocation under arbitrary varying channels: an online approach. *IEEE/ACM Transactions on Networking (TON)*, 20(2):477–487, 2012.

Patrick Busch and Richa Malhotra. Wireless lan with channel swapping between dfs access points, April 3 2012. US Patent 8,150,955.

Murat ÇakiroÇğlu and Ahmet Turan Özcerit. Jamming detection mechanisms for wireless sensor networks. In *Proceedings of the 3rd international conference on Scalable information systems*, page 4. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008.

Zenon Chaczko, Ruckshan Wickramasooriya, Ryszard Klempous, and Jan Nikodem. Security threats in cognitive radio applications. In *Intelligent Engineering Systems (INES), 2010 14th International Conference on*, pages 209–214. IEEE, 2010.

Zenon Chaczko, Shaher Slehar, and Tamer Shnoudi. Game-theory based cognitive radio policies for jamming and anti-jamming in the iot. In *2018 12th International Symposium on Medical Information and Communication Technology (ISMICT)*, pages 1–6. IEEE, 2018.

Hsi-Lu Chao, Tzung-Lin Li, Cheng-Che Chung, and Sau-Hsuan Wu. Throughput analysis of a hybrid mac protocol for wifi-based heterogeneous cognitive radio networks. In *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, pages 1–5. IEEE, 2015.

Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In *Advances in neural information processing systems*, pages 2249–2257, 2011.

K-C Chen, Y-J Peng, Neeli Prasad, Y-C Liang, and Sumei Sun. Cognitive radio network architecture: part i-general structure. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pages 114–119. ACM, 2008.

Ruiliang Chen and Jung-Min Park. Ensuring trustworthy spectrum sensing in cognitive radio networks. In *Networking Technologies for Software Defined Radio Networks, 2006. SDR'06.1 st IEEE Workshop on*, pages 110–119. IEEE, 2006.

Kresimir Dabcevic. Intelligent jamming and anti-jamming techniques using cognitive radios. *PhD Programme in Computational Intelligence University of Genoa*, 2015.

Luca De Nardis and Oliver Holland. Deployment scenarios for cognitive radio. In *Cognitive Radio Policy and Regulation*, pages 49–116. Springer, 2014.

J Josephine Dhivya and M Ramaswami. Analysis of handoff parameters in cognitive radio networks on coadunation of wifi and wimax systems. In *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, pages 190–194. IEEE, 2017.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Pac bounds for multi-armed bandit and markov decision processes. In *International Conference on Computational Learning Theory*, pages 255–270. Springer, 2002.

Sarah Filippi. Optimistic strategies in reinforcement learning. *Theses, Ecole nationale supérieure des telecommunications-ENST*, 134, 2010.

National Science Foundation. Network simulator 3, 2016. URL `http://www.nsnam.org`.

Alexandros G Fragkiadakis, Elias Z Tragos, and Ioannis G Askoxylakis. A survey on security threats and detection techniques in cognitive radio networks. *Communications Surveys & Tutorials, IEEE*, 15(1):428–445, 2013.

Aurélien Garivier and Olivier Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual Conference On Learning Theory*, pages 359–376, 2011.

Andrey Garnaev, Yezekael Hayel, and Eitan Altman. A bayesian jamming game in an ofdm wireless network. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2012 10th International Symposium on*, pages 41–48. IEEE, 2012.

Sudhanshu Gaur, Jeng-Shiann Jiang, Mary Ann Ingram, and M Fatih Demirkol. Interfering mimo links with stream control and optimal antenna selection. In *Global Telecommunications Conference, 2004. GLOBECOM'04. IEEE*, volume 5, pages 3138–3142. IEEE, 2004.

John C Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 148–177, 1979.

Cognitive Radio Working Group et al. Sdrf cognitive radio definitions (sdrf-06-r-0011-v1. 0.0). Technical report, The Wireless Innovation Forum, Retrieved on September 24 2012, from URL http://âĂŃ groups.âĂŃ winnforum.âĂŃ org/âĂŃ d/âĂŃ do/âĂŃ 1585, retrieved on September 24, 2012, from http://âĂŃ groups.âĂŃ winnforum.âĂŃ org/âĂŃ d/âĂŃ do/âĂŃ 1585, 2007.

Kanika Grover, Alvin Lim, and Qing Yang. Jamming and anti–jamming techniques in wireless networks: a survey. *International Journal of Ad Hoc and Ubiquitous Computing*, 17(4):197–215, 2014.

Luis Guijarro, Vicent Pla, and Jose R Vidal. Competition in cognitive radio networks: spectrum leasing and innovation. In *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, pages 1112–1113. IEEE, 2011.

Youngjune Gwon, Siamak Dastangoo, and HT Kung. Optimizing media access strategy for competing cognitive radio networks. In *2013 IEEE Global Communications Conference (GLOBECOM)*, pages 1215–1220. IEEE, 2013.

Hiroshi Harada. Software defined radio prototype toward cognitive radio communication systems. In *New Frontiers in Dynamic Spectrum Access Networks, 2005.*

DySPAN 2005. 2005 First IEEE International Symposium on, pages 539–547. IEEE, 2005.

Hiroshi Harada. Research and development on cognitive and software radio technologies-devices and hardware platform. *General assembly of URSI*, 2008.

Juan Hernandez-Serrano, Olga León, and Miguel Soriano. Modeling the lion attack in cognitive radio networks. *EURASIP Journal on Wireless Communications and Networking*, 2011:2, 2011.

Dinh Thai Hoang, Dusit Niyato, Ping Wang, Dong In Kim, and Zhu Han. Ambient backscatter: A new approach to improve network performance for rf-powered cognitive radio networks. *IEEE Transactions on Communications*, 65(9):3659–3674, 2017.

Yih-Chun Hu, David B Johnson, and Adrian Perrig. Sead: Secure efficient distance vector routing for mobile wireless ad hoc networks. *Ad hoc networks*, 1(1):175–192, 2003.

Furong Huang, Wei Wang, Haiyan Luo, Guanding Yu, and Zhaoyang Zhang. Prediction-based spectrum aggregation with hardware limitation in cognitive radio networks. In *2010 IEEE 71st Vehicular Technology Conference*, pages 1–5. IEEE, 2010.

Hanen Idoudi, Kevin Daimi, and Mustafa Saed. Security challenges in cognitive radio networks. In *Proceedings of the World Congress on Engineering*, volume 1, 2014.

Kentaro Ishizu and Hiroshi Harada. A load-balancing framework for cognitive wireless network to coexist with legacy wifi systems. In *VTC Spring 2009-IEEE 69th Vehicular Technology Conference*, pages 1–5. IEEE, 2009.

Wang Jinlong, Feng Shuo, Wu Qihui, Zheng Xueqiang, and Xu Yuhua. Hierarchical cognition cycle for cognitive radio networks. *China Communications*, 12(1):108–121, 2015.

Aleksandar Jovicic and Pramod Viswanath. Cognitive radio: An information-theoretic perspective. *IEEE Transactions on Information Theory*, 55(9):3945–3958, 2009.

K Karunambiga, AC Sumathi, and M Sundarambal. Channel selection strategy for jamming-resistant reactive frequency hopping in cognitive wifi network. In *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*, pages 1–4. IEEE, 2015.

Michael N Katehakis and Herbert Robbins. Sequential choice from several populations. *Proceedings of the National Academy of Sciences of the United States of America*, 92(19):8584, 1995.

Ahmed Khattab, Dmitri Perkins, and Magdy Bayoumi. Cognitive radio networking preliminaries. In *Cognitive Radio Networks*, pages 11–20. Springer, 2013.

George W Kibirige and Camilius Sanga. A survey on detection of sinkhole attack in wireless sensor network. *arXiv preprint arXiv:1505.01941*, 2015.

Levent Koçkesen and Efe A Ok. An introduction to game theory. *University Efe A. Ok New York University July*, 8, 2007.

P Kolodzy et al. Next generation communications: Kickoff meeting. In *Proc. DARPA*, volume 10, 2001.

Nathaniel Korda, Emilie Kaufmann, and Remi Munos. Thompson sampling for 1-dimensional exponential family bandits. In *Advances in Neural Information Processing Systems*, pages 1448–1456, 2013.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

Tor Lattimore and Csaba Szepesvári. Bandit algorithms. *preprint*, 2018.

Che Wei Lee, Show-Shiow Tzeng, and Ying-Jen Lin. Energy-efficient transmission scheme for vehicles with cognitive radio/wifi in vehicular networks. In *2015 IEEE International Conference on Data Science and Data Intensive Systems*, pages 470–471. IEEE, 2015.

Xiaohua Li and Wednel Cadeau. Anti-jamming performance of cognitive radio networks. In *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, pages 1–6. IEEE, 2011.

Ying Chang Liang, Kwang Cheng Chen, Geoffrey Ye Li, and Petri 2011 Mähönen. Cognitive radio networking and communications: An overview. *IEEE Transactions on Vehicular Technology*, 60(7):3386–3407, 2011. ISSN 00189545. doi: 10.1109/TVT.2011.2158673.

Mee Hong Ling, Kok-Lim Alvin Yau, Junaid Qadir, Geong Sen Poh, and Qiang Ni. Application of reinforcement learning for security enhancement in cognitive radio networks. *Applied Soft Computing*, 37:809–829, 2015.

Yao Liu and Peng Ning. Bittrickle: Defending against broadband and high-power reactive jamming attacks. In *INFOCOM, 2012 Proceedings IEEE*, pages 909–917. IEEE, 2012.

Stephen Machuzak and Sudharman K Jayaweera. Reinforcement learning based anti-jamming with wideband autonomous cognitive radios. In *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, pages 1–5. IEEE, 2016.

A Hyils Sharon Magdalene and L Thulasimani. Analysis of spectrum sensing data falsification (ssdf) attack in cognitive radio networks: A survey. *Journal of Science & Engineering Education (ISSN 2455-5061)*, 2:89–100, 2017.

Maheswari Mangai, Shanmuga Sundaram, Needra Fernando, Vijay Daniel, and Suresh Babu. A state of the art review on various security threats in cognitive radio networks. *International Journal of Computer Science and Mobile Computing*, 2(12):128–144, 2013.

C Manogna and K Bhaskar Naik. Detection of Jamming Attack in Cognitive Radio Networks 1 1. pages 69–72, 2014.

Huaqing Mao and Li Zhu. An investigation on security of cognitive radio networks. In *Management and Service Science (MASS), 2011 International Conference on*, pages 1–4. IEEE, 2011.

James D Mccaffrey. The Epsilon-Greedy Algorithm. 2018.

Roger McFarlane. A survey of exploration strategies in reinforcement learning. *McGill University, http://www. cs. mcgill. ca/ cs526/roger. pdf, accessed: April*, 2018.

Andrew R McGee, Matthieu Coutière, and Maria E Palamara. Public safety network security considerations. *Bell Labs Technical Journal*, 17(3):79–86, 2012.

Rahul Meshram, D Manjunath, and Aditya Gopalan. On the whittle index for restless multiarmed hidden markov bandits. *IEEE Transactions on Automatic Control*, 63(9):3046–3053, 2018.

Joseph Mitola. Cognitive radio-an integrated agent architecture for software defined radio. 2000.

Joseph Mitola. *Cognitive Radio Architecture: The Engineering Foundations of Radio XML*. 2005. ISBN 0471742449. doi: 10.1002/0471773735.

Joseph Mitola Iii. Cognitive radio for flexible mobile multimedia communications. In *Mobile Multimedia Communications, 1999.(MoMuC'99) 1999 IEEE International Workshop on*, pages 3–10. IEEE, 1999.

Radovan Miucic. *Connected Vehicles: Intelligent Transportation Systems*. Springer, 2018.

Apurva N Mody, Ranga Reddy, Thomas Kiernan, and Timothy X Brown. Security in cognitive radio networks: An example using the commercial ieee 802.22 standard. In *Military Communications Conference, 2009. MILCOM 2009. IEEE*, pages 1–7. IEEE, 2009.

Pedro J Moreno, Purdy P Ho, and Nuno Vasconcelos. A kullback-leibler divergence based kernel for svm classification in multimedia applications. In *Advances in neural information processing systems*, pages 1385–1392, 2004.

S Bhagavathy Nanthini, M Hemalatha, D Manivannan, and L Devasena. Attacks in cognitive radio networks (crn)-a survey. *Indian Journal of Science and Technology*, 7(4):530, 2014.

Maziar Nekovee. Cognitive radio access to tv white spaces: Spectrum opportunities, commercial applications and remaining technology challenges. In *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*, pages 1–10. IEEE, 2010.

IEEEWorking Group onWireless Regional Area Networks. Enabling rural broadband wireless access using cognitive radio technology in TV whitespaces., 2014. URL `http://www.ieee802.org/22/`.

Danh Nguyen, Cem Sahin, Boris Shishkin, Nagarajan Kandasamy, and Kapil R Dandekar. A real-time and protocol-aware reactive jamming framework built on software-defined radios. In *Proceedings of the 2014 ACM workshop on Software radio implementation forum*, pages 15–22. ACM, 2014.

Van Tam Nguyen, Frederic Villain, and Yann Le Guillou. Cognitive radio rf: overview and challenges. *VLSI Design*, 2012:1, 2012.

André Oliveira, Zhili Sun, Philippe Boutry, Diego Gimenez, Antonio Pietrabissa, and Katja Banovec Juros. Internetworking and wireless ad hoc networks for emergency and disaster relief services. *International Journal of Satellite Communications Policy and Management*, 1(1):1–14, 2011.

Tayfun Ozdemir. A polarization diversity multi-beam antenna for zigbee applications. In *2009 IEEE 10th Annual Wireless and Microwave Technology Conference*, pages 1–4. IEEE, 2009.

Sazia Parvin and Farookh Khadeer Hussain. Digital signature-based secure communication in cognitive radio networks. In *Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on*, pages 230–235. IEEE, 2011.

Qing-Qi Pei, Hong-Ning Li, Hong-Yang Zhao, Nan Li, and Ying Min. Security in cognitive radio networks. *Journal of China Institute of Communications*, 34(1): 144–158, 2013.

Yiyang Pei, Ying-Chang Liang, Lan Zhang, Kah Chan Teh, and Kwok Hung Li. Secure communication over miso cognitive radio channels. *Wireless Communications, IEEE Transactions on*, 9(4):1494–1502, 2010.

Sudeep Raja. Multi armed bandits and exploration strategies. *Multi Armed Bandits and Exploration Strategies–Sudeep Raja–MS/Phd Student at UMass Amherst*, 28, 2016.

Ram Ramanathan and Criag Partridge. Next generation (xg) architecture and protocol development (xap). Technical report, DTIC Document, 2005.

R Ramesh, J Jerald, Tom Page, and Subramaniam Arunachalam. Concurrent tolerance allocation using an artificial neural network and continuous ant colony optimisation. *International Journal of Design Engineering*, 2(1):1–25, 2009.

Priyanka Rawat, Kamal Deep Singh, and Jean Marie Bonnin. Cognitive radio for m2m and internet of things: A survey. *Computer Communications*, 94:1–29, 2016.

Riverbed Technology. Opnet, 2016. URL `http://www.riverbed.com/products/performance-management-control/opnet.html`.

Ronald L Rivest and Yiqun Yin. Simulation results for a new two-armed bandit heuristic. In *Proceedings of a workshop on Computational learning theory and natural learning systems (vol. 1): constraints and prospects: constraints and prospects*, pages 477–486. MIT Press, 1994.

Herbert Robbins. Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers*, pages 169–177. Springer, 1985.

Daniel Russo, Benjamin Van Roy, Abbas Kazerouni, and Ian Osband. A tutorial on thompson sampling. *arXiv preprint arXiv:1707.02038*, 2017.

Ghazanfar Ali Safdar and MO Neill. Common control channel security framework for cognitive radio networks. In *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pages 1–5. IEEE, 2009.

Ashwin Sampath, Hui Dai, Haitao Zheng, and Ben Y Zhao. Multi-channel jamming attacks using cognitive radios. In *Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference on*, pages 352–357. IEEE, 2007.

Steven L Scott. A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26(6):639–658, 2010.

Stephen J Shellhammer, Ahmed K Sadek, and Wenyi Zhang. Technical challenges for cognitive radio in the tv white space spectrum. In *2009 Information Theory and Applications Workshop*, pages 323–333. IEEE, 2009.

Simulcraft Inc. Omnest, 2015. URL `http://www.omnest.com`.

Feten Slimeni, Bart Scheers, Zied Chtourou, Vincent Le Nir, and Rabah Attia. Cognitive radio jamming mitigation using markov decision process and reinforcement learning. *Procedia Computer Science*, 73:199–208, 2015.

K Sridhara, Ashok Chandra, and Purnendu SM Tripathi. Spectrum challenges and solutions by cognitive radio: An overview. *Wireless Personal Communications*, 45(3):281–291, 2008.

Peter Steenkiste, Douglas Sicker, Gary Minden, and Dipankar Raychaudhuri. Future directions in cognitive radio network research. In *NSF workshop report*, volume 4, pages 1–2, 2009.

John M Stewart. *Python for scientists*. Cambridge University Press, 2017.

Malcolm Strens. A bayesian framework for reinforcement learning. In *ICML*, pages 943–950, 2000.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

Viktor Toldov, Laurent Clavier, Valeria Loscrí, and Nathalie Mitton. A thompson sampling approach to channel exploration-exploitation problem in multihop cognitive radio networks. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2016 IEEE 27th Annual International Symposium on*, pages 1–6. IEEE, 2016.

Ivana Tomić and Julie A McCann. A survey of potential security issues in existing wireless sensor network protocols. *IEEE Internet of Things Journal*, 4(6):1910–1923, 2017.

ICT Regulation Toolkit. What is spectrum Trading, 2010.

Lloyd N Trefethen. *Spectral methods in MATLAB*, volume 10. Siam, 2000.

Shahnaza Tursunova, Khamidulla Inoyatov, and Young-Tak Kim. Cognitive passive estimation of available bandwidth (cpeab) in overlapped ieee 802.11 wifi wlans. In *2010 IEEE Network Operations and Management Symposium-NOMS 2010*, pages 448–454. IEEE, 2010.

Andras Varga. Omnet++. In *Modeling and tools for network simulation*, pages 35–59. Springer, 2010.

Abraham Wald. *Sequential analysis*. Courier Corporation, 1973.

Beibei Wang, Yongle Wu, and KJ Ray Liu. Game theory for cognitive radio networks: An overview. *Computer networks*, 54(14):2537–2561, 2010a.

Beibei Wang, Yongle Wu, KJ Liu, and T Charles Clancy. An anti-jamming stochastic game for cognitive radio networks. *Selected Areas in Communications, IEEE Journal on*, 29(4):877–889, 2011.

Le Wang and Alexander M Wyglinski. A combined approach for distinguishing different types of jamming attacks against wireless networks. In *Communications, Computers and Signal Processing (PacRim), 2011 IEEE Pacific Rim Conference on*, pages 809–814. IEEE, 2011.

Wenkai Wang, Husheng Li, Yan Lindsay Sun, and Zhu Han. Attack-proof collaborative spectrum sensing in cognitive radio networks. In *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pages 130–134. IEEE, 2009.

Wenkai Wang, Yan Sun, Husheng Li, and Zhu Han. Cross-layer attack and defense in cognitive radio networks. In *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, pages 1–6. IEEE, 2010b.

Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, pages 287–298, 1988.

Matthias Wilhelm, Ivan Martinovic, Jens B Schmitt, and Vincent Lenders. Short paper: reactive jamming in wireless networks: how realistic is the threat? In *Proceedings of the fourth ACM conference on Wireless network security*, pages 47–52. ACM, 2011.

Jeremy Wyatt. Exploration and inference in learning from reinforcement. 1998.

Wenyuan Xu, Timothy Wood, Wade Trappe, and Yanyong Zhang. Channel surfing and spatial retreats: defenses against wireless denial of service. In *Proceedings of the 3rd ACM workshop on Wireless security*, pages 80–89. ACM, 2004.

Wenyuan Xu, Wade Trappe, Yanyong Zhang, and Timothy Wood. The feasibility of launching and detecting jamming attacks in wireless networks. In *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, pages 46–57. ACM, 2005.

Tevfik Yücek and Hüseyin Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *Communications Surveys & Tutorials, IEEE*, 11(1): 116–130, 2009.

Linyuan Zhang, Guoru Ding, Qihui Wu, Yulong Zou, Zhu Han, and Jinlong Wang. Byzantine attack and defense in cognitive radio networks: A survey. *IEEE Communications Surveys & Tutorials*, 17(3):1342–1363, 2015.

Fuhui Zhou, Yongpeng Wu, Ying-Chang Liang, Zan Li, Yuhao Wang, and Kai-Kit Wong. State of the art, taxonomy, and open issues on cognitive radio networks with noma. *IEEE Wireless Communications*, 25(2):100–108, 2018.

Li Zhu and Huaibei Zhou. Two types of attacks against cognitive radio network mac protocols. In *Computer Science and Software Engineering, 2008 International Conference on*, volume 4, pages 1110–1113. IEEE, 2008.